

A new world ocean temperature profile product

The Internationally Q_uality- Controlled Ocean Database v0.1

(IQuOD v0.1)



GSOP



SG-IQuOD



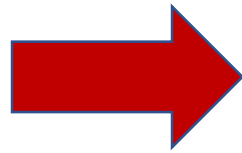
WG 148

Outline

- **Background: Who is IQuOD? What, Why, How**
- **First interim product release: IQuOD v0.1**
- **What is in the pipeline?**
- **Invitation to join us**

Quality (meet requirements)

Data



Knowledge

Subsurface Ocean
Temperature

EOV & ECV

Challenges

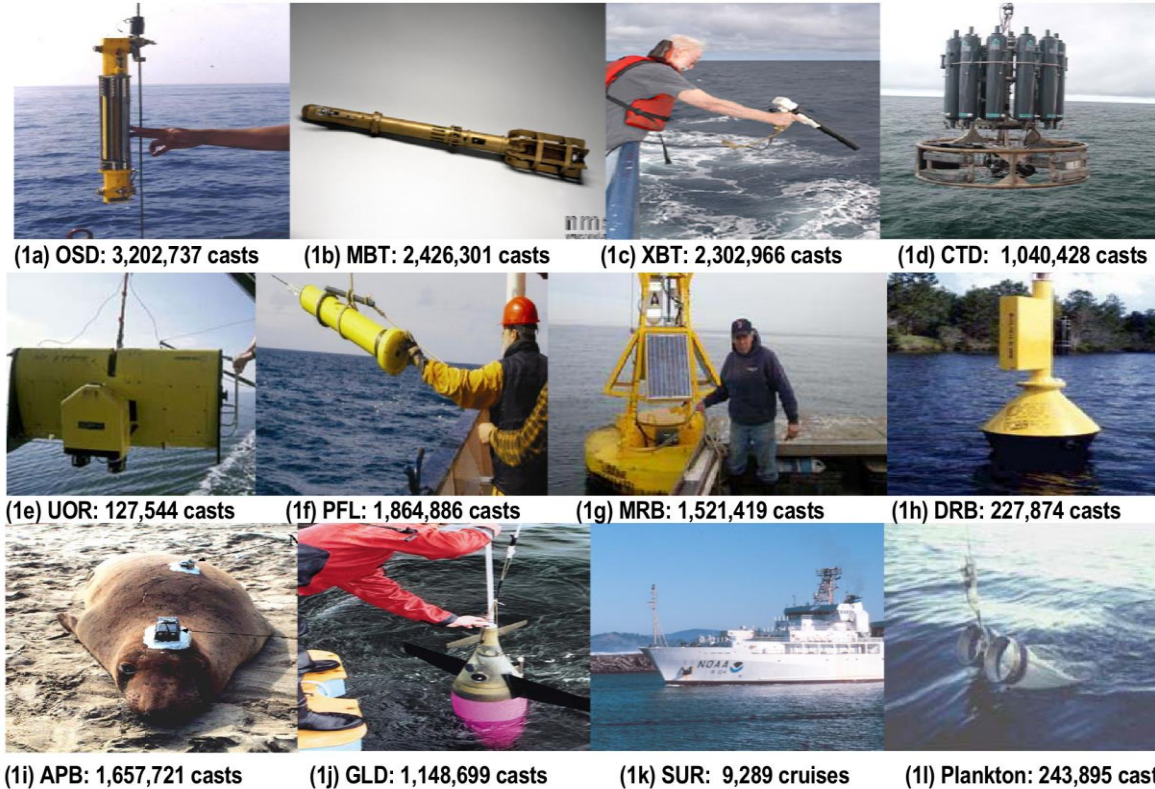
Society: climate change

Observ.: most stringent quality

- Weather and climate prediction services
- Ocean climate variability and change
(timeline + natural/human induced)
- Earth's energy, sea level and water budgets
- Constrain climate projections
- Adaptation and mitigation (policymaking)

World Ocean Database:

World's largest publicly available oceanographic profile database



Courtesy:
Tim Boyer

**Highly heterogeneous temperature obs. taken for various purposes, not all climate quality.
No coordination: not necessarily the most efficient QC methods & duplication of efforts**

IQuOD in a nutshell

To **maximize the quality, consistency and completeness** of the long-term world's ocean **subsurface temperature** database subsurface profiles | (intelligent) metadata | uncertainty

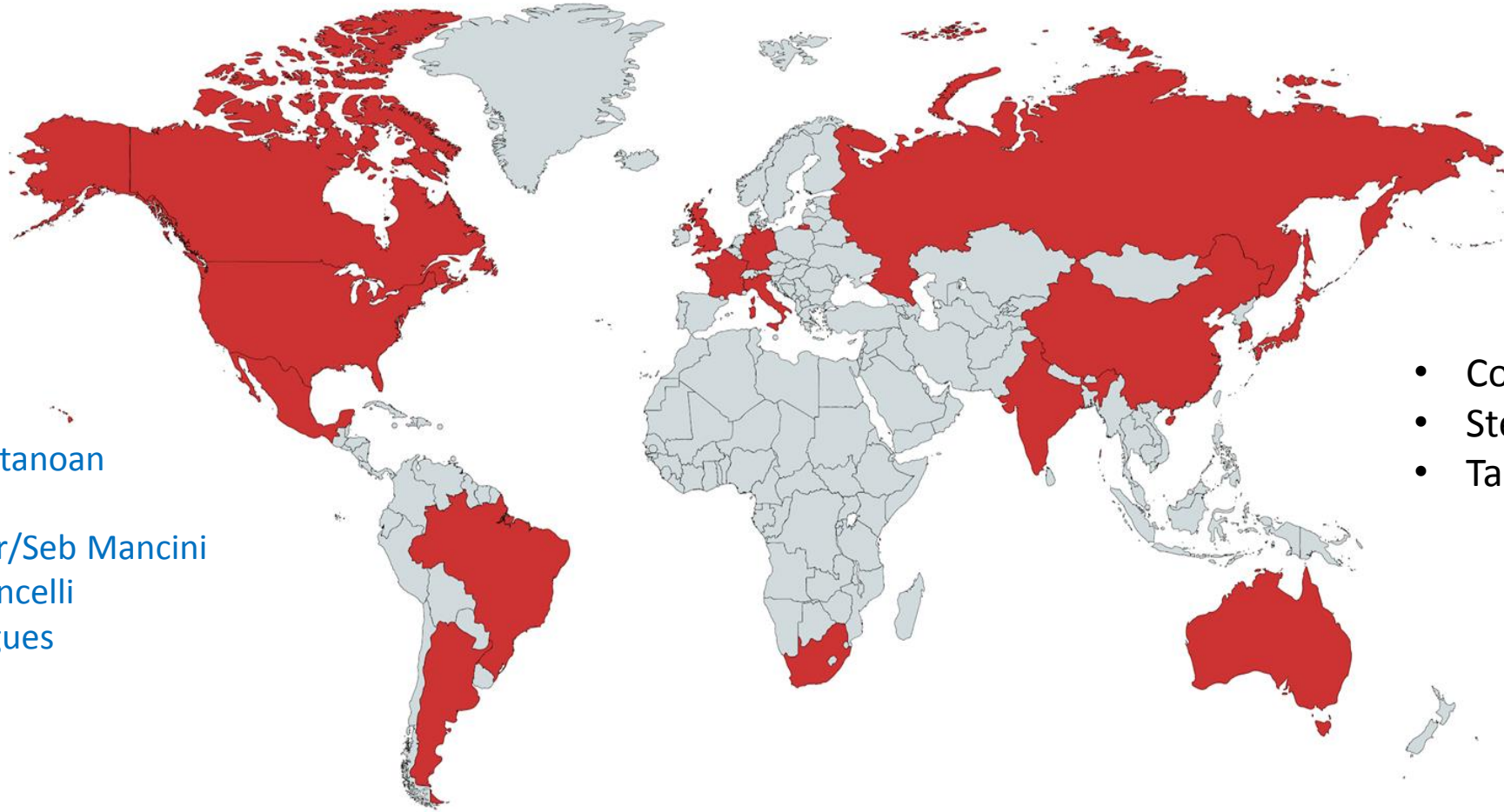
Through **international coordination** of resources and expertise into a **single best practice community effort**.

Supported by:

- IQuOD is a high priority activity of the Global Panel (GSOP): [WCRP CLIVAR Science Plan for 2015-2025](#)
- The [IODE-IQuOD](#) project was established by IODE-XXIII (2015) through [Recommendation IODE-XXIII.3](#)
- SCOR WG 148 IQuOD: <http://www.scor-int.org/Annual%20Meetings/2015EC/IQuOD.pdf>



IQuOD team: globally diverse (17 nation members)



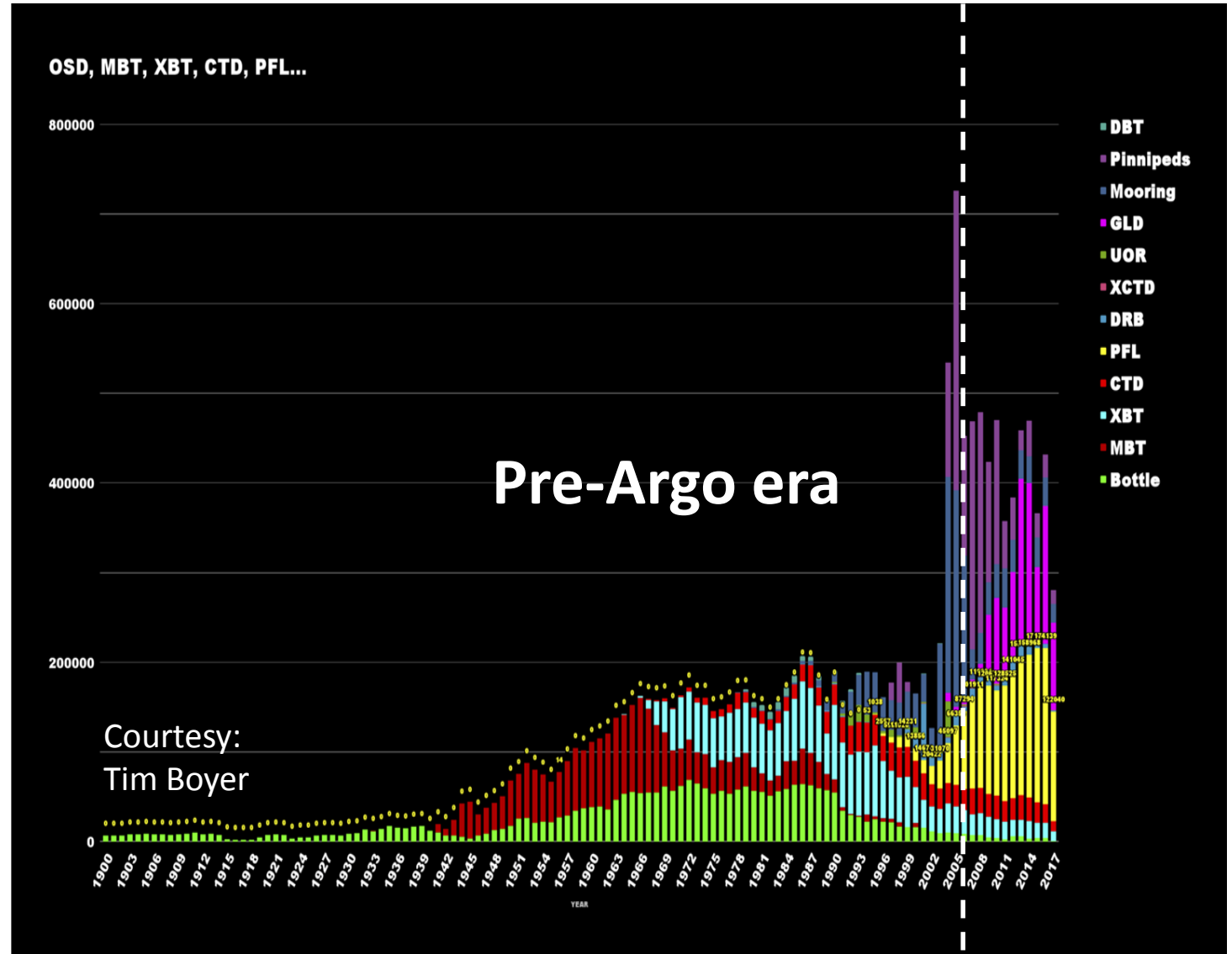
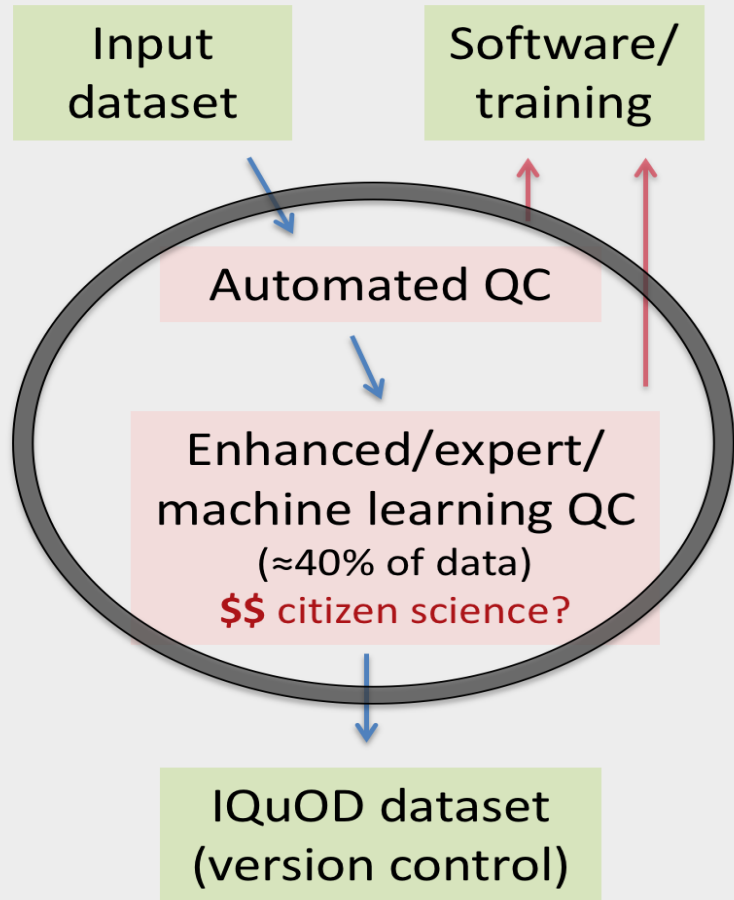
At IMDIS:

- Toru Suzuki
- Christine Coatanoan
- Loic Villion
- Roger Proctor/Seb Mancini
- Simona Simoncelli
- Catia Domingues

- Co-chairs
- Steering committee
- Task teams

Close collaboration among experts and users (data quality and management, data assimilation, modelers and the broader climate community)

Workflow



WORLD OCEAN DATABASE (n.obs/ year)

IQuOD v0.1: released 16th March 2018

- ✓ **I-metadata**: “first cut” algorithms for unknown XBTs, building on from Cowley et al. (2013) and recoded into open-source Python (Palmer et al. 2018, J. Atm. Oc. Tech.)
- ✓ **Uncertainty**: “first cut” random error attached to each discrete observation (Cowley et al. draft in preparation)
- ✓ **Format**: ASCII and netCDF ragged array (CF compliant)
- ✓ **GDAC**: distribution via US NCEI
(in progress: UK MetOffice, Japan ODC, French Coriolis, Chinese ODC and Australian IMOS, SOOSmap/EMODNET (via Pip Bricher))



International Quality Controlled Ocean Database (IQuOD) version 0.1 - aggregated and community quality controlled ocean profile data 1772-present



Preview graphic

This data set includes subsurface ocean profiles of temperature, salinity, oxygen, nutrients, ocean tracers, optics, and biology (chlorophyll, plankton) taken from 1772 to 2018 in the global ocean using bottles, CTD, XBT, MBT, profiling floats, moored buoys, ice drifting buoys, gliders, towed profilers, and instrumented pinnipeds. This data set was prepared at NCEI in CF compliant netCDF ragged array format under the direction of the IQuOD project. The IQuOD (International Quality-controlled Ocean Database) effort is being organized by the oceanographic community, and includes experts in data quality and management, climate modelers and the broader climate-related community. The primary focus of IQuOD is to produce and freely distribute the highest quality and complete single ocean profile repository.

[Dataset Citation](#)

[Dataset Identifiers](#)

[ISO 19115-2 Metadata](#)

Access	Time & Location	Documentation	Description	Credit	Keywords	Constraints	Lineage
General Documentation		<p>IQuOD Version 0.1 - 16 March 2018</p> <p>Navigate directly to the URL for a descriptive web page.</p> <p>Dataset Description</p> <p>Navigate directly to the URL for a descriptive web page with download links.</p>					
Associated Resources		<p>International Quality controlled Ocean Database (IQuOD)</p> <ul style="list-style-type: none"> IQuOD project web site 					

Last Modified: 2018-10-24T22:43:23

For questions about the information on this page, please email: NODC.DataOfficer@noaa.gov

IQuOD v0.1

Access	Time & Location	Documentation	Description	Credit	Keywords	Constraints	Lineage
Download Data		<p>THREDDS (download)</p> <p>These data are available through a variety of services via a THREDDS (Thematic Real-time Environmental Distributed Data Services) Data Server (TDS). Depending on the dataset, the TDS can provide WMS, WCS, DAP, HTTP, and other data access and metadata services as well. For more information on the TDS, see http://www.unidata.ucar.edu/software/thredds/current/tds/.</p> <p>HTTP (download)</p> <p>Navigate directly to the URL for data access and direct download.</p> <p>FTP (download)</p> <p>These data are available through the File Transfer Protocol (FTP). You may use any FTP client to download these data.</p>					
Distribution Formats		<ul style="list-style-type: none"> netCDF (Version: netCDF-4) <ul style="list-style-type: none"> File Specification: IQuOD Multi-cast file 					
Ordering Instructions		<p>Data may be searched and downloaded using online services provided by NCEI using the online resource URLs in this record. Contact NCEI Information Services for custom orders. When requesting data from NCEI, the desired data set may be referred to by the unique package identification number listed in this metadata record.</p>					
Distributor		<p>DOC/NOAA/NESDIS/NCEI > National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce</p> <p>301-713-3277</p> <p>NCEI.Info@noaa.gov</p>					
Dataset Point of Contact		<p>Information Services</p> <p>DOC/NOAA/NESDIS/NCEI > National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce</p> <p>301-713-3277</p> <p>NCEI.Info@noaa.gov</p>					

Last Modified: 2018-10-24T22:43:23

For questions about the information on this page, please email: NODC.DataOfficer@noaa.gov

What is in the pipeline?

Tasks in progress/to be implemented

- ✓ **Format**: ASCII and netCDF ragged array (CF compliant). Implementation of feedback from users (data assimilation community)
- ✓ **Duplicates**: exact duplicates flagged. Implementation of Guillaume Maze's machine learning algorithm for detection of duplicates.
- ✓ **I-metadata**: moving from a deterministic to a probabilistic/neural networks approaches for XBTs
- ✓ **Uncertainty**: refinements in random error assessments

I-metadata XBT: deterministic approach

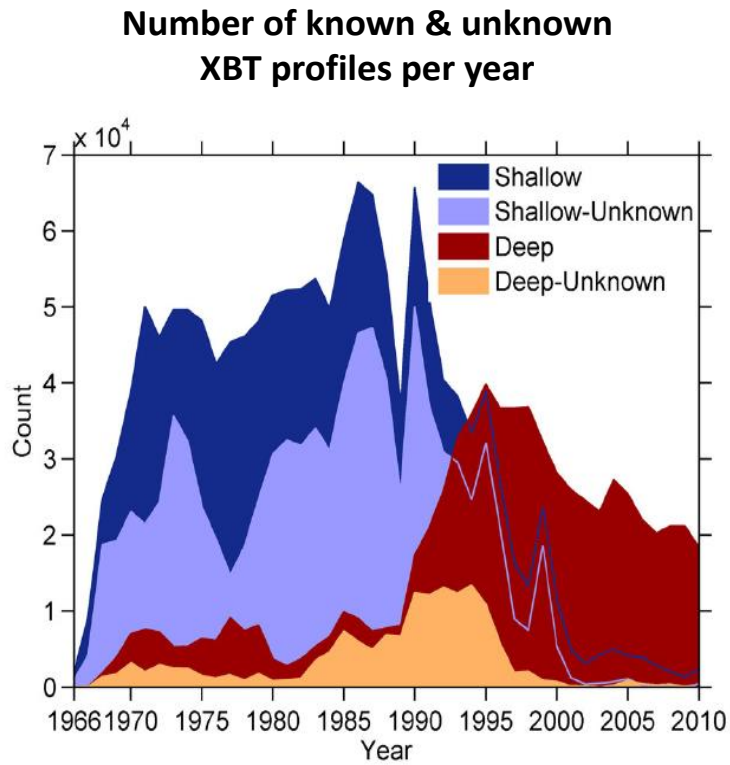
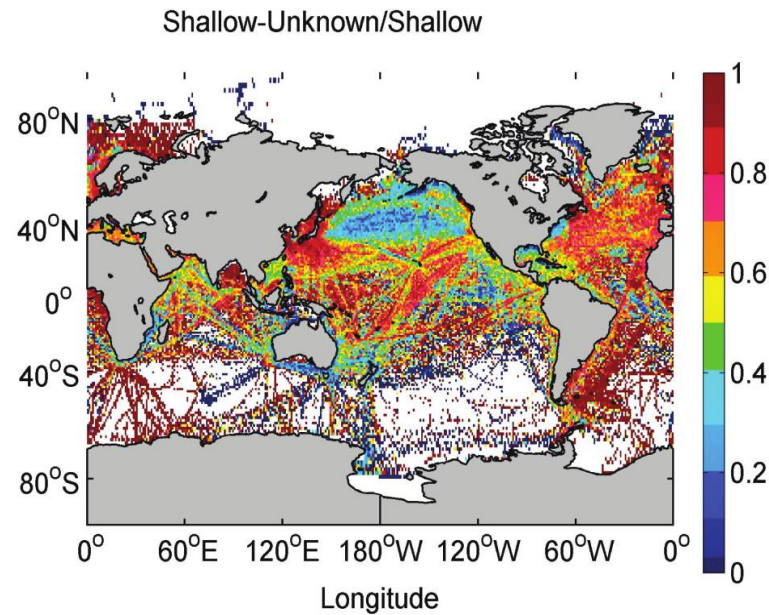
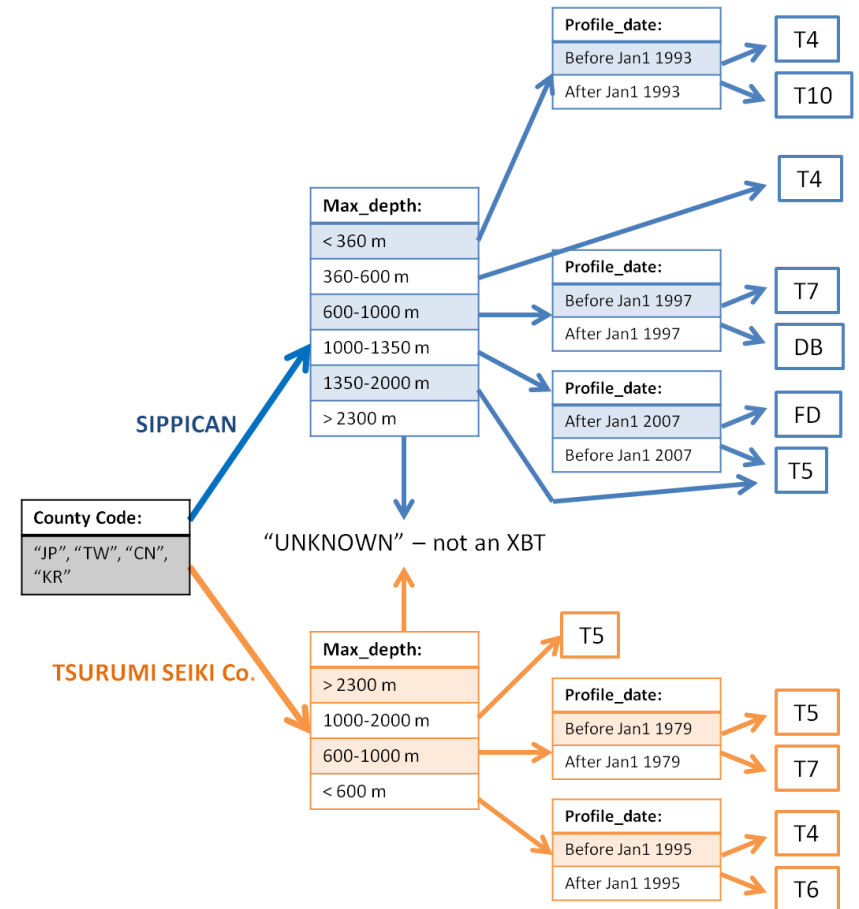


Figure 2. Total number of shallow (dark blue) and deep (dark red) XBT profiles per year and the number of these for which the type is unknown (shallow = light blue; deep = orange).

Proportion of unknown XBT profiles (shallow)



Abraham et al. (2013)



Palmer et al. 2018



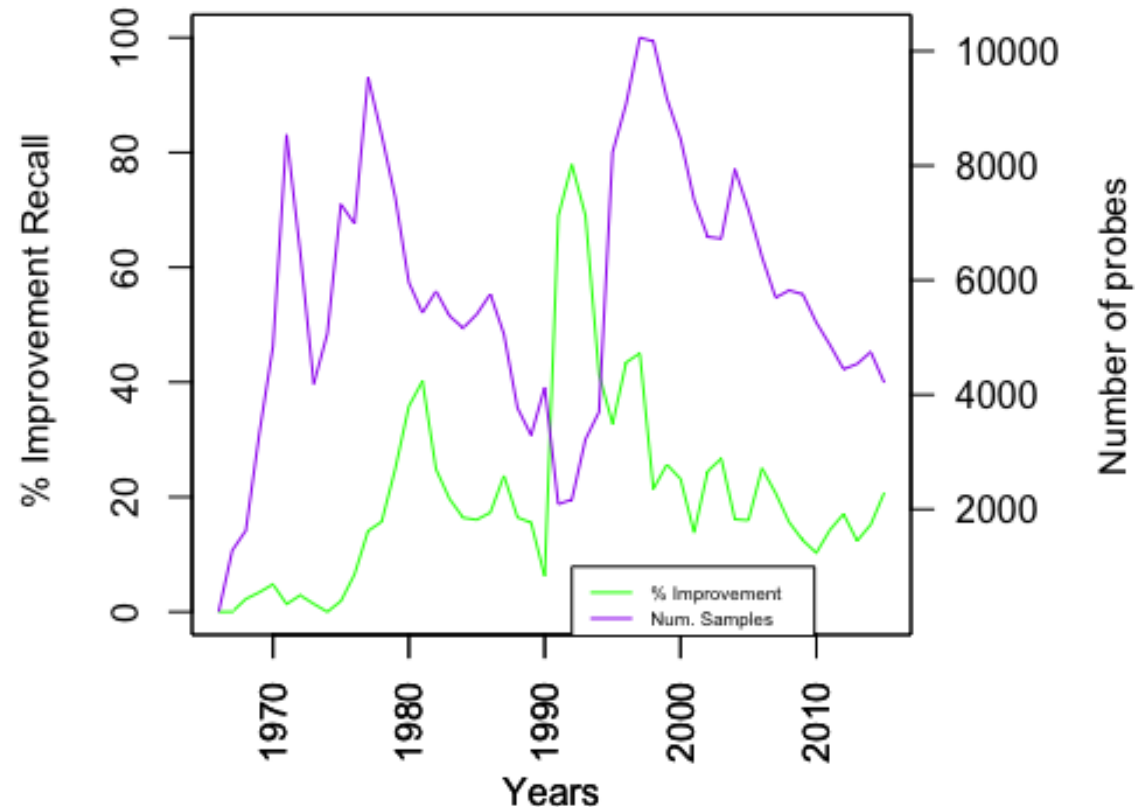
<http://www.informaticslab.co.uk>

“Using Neural Networks to Correct Historical Climate Observations”

T.P. Leahy, F. Pons Llopis, M.D. Palmer and N.H. Robinson

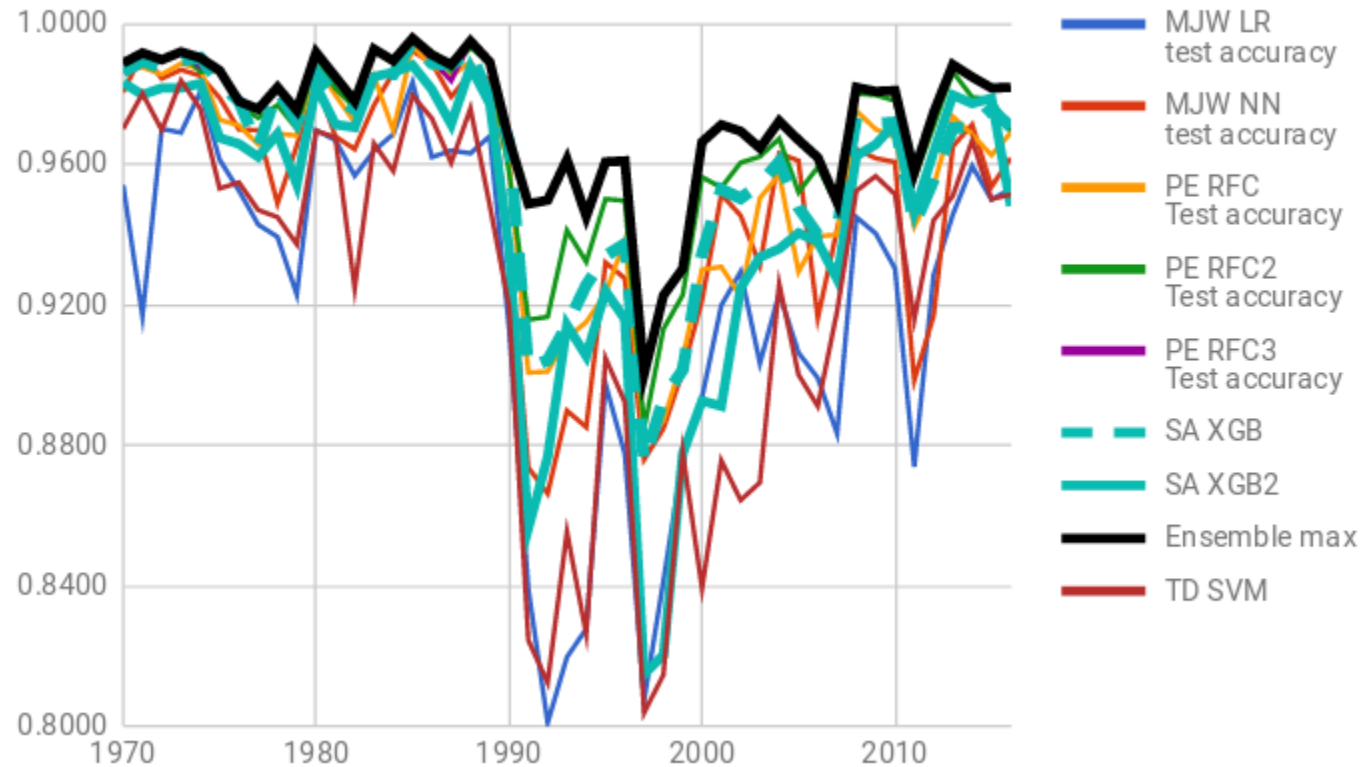
submitted to *Journal of Atmospheric and Oceanic Technology*

- Substantial improvements on v0.1 iMeta algorithm
- **However:** more in-depth work needed on training and validation datasets
- Potential for **over-confidence** in classification of unknown probes



Comparison of Machine Learning

XBT test accuracy (Rachel's split)



- ALL machine learning methods yield an accuracy of 80% or better
- Even better accuracy may be possible using an ensemble approach

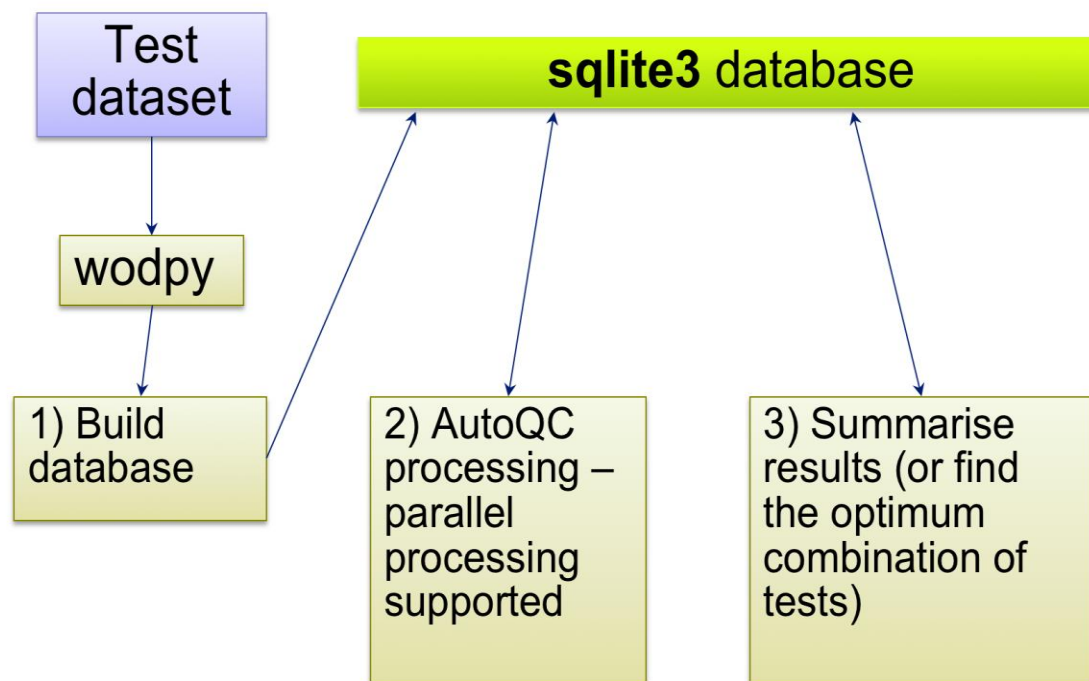
I-metadata where to next?

- More discussion / insights into emergent relationships in the ML methods
- Which predictor variables add skill? How to best formulate the “cost function”?
- Development of probe type probabilities to generate rate a Monte Carlo of probe + manufacturer assignment
- **Investigate the impact of iMeta Data assignments on estimates of historical ocean heating rates**



The AutoQC processor

Finding the set of most effective automatic quality control tests



<https://github.com/IQuOD>

<https://github.com/IQuOD>

WodPy

AutoQC

General purpose Python reader for WOD ASCII format data

Our QC code and benchmarking software

Also available as a docker image:

To install:
pip install wodpy

iquod/autoqc

www.metoffice.gov.uk

Updated: 2018.04.16

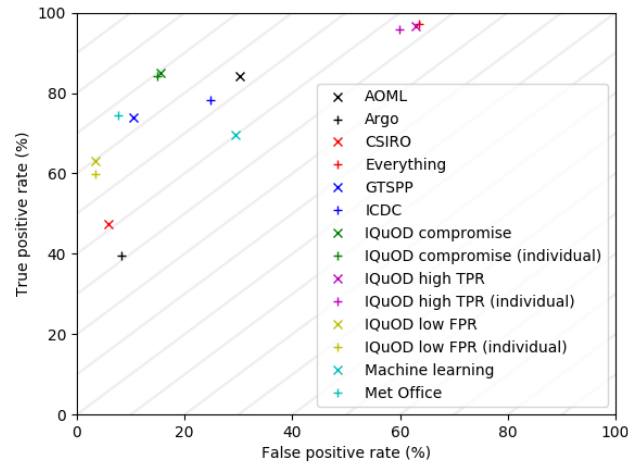
© Crown Copyright 2016, Met Office

GitHub, Inc. [US] | <https://github.com/IQuOD/AutoQC>

The screenshot shows the GitHub repository page for IQuOD/AutoQC. The repository has 773 commits, 13 branches, 3 releases, and 6 contributors. The commit history is visible, showing recent updates to various files like 'cotede_qc', 'data', 'qctests', 'tests', 'util', '.gitignore', '.travis.yml', 'AutoQC.py', 'Dockerfile', 'LICENSE', 'README.md', 'analyse-results.py', 'build-db.py', 'catchall.py', 'filter-db.py', 'install.sh', 'plot-roc.py', 'qctest_groups.csv', 'qctest_requirements.json', and 'summarize-results.py'.

Benchmarking: 55 AQC tests

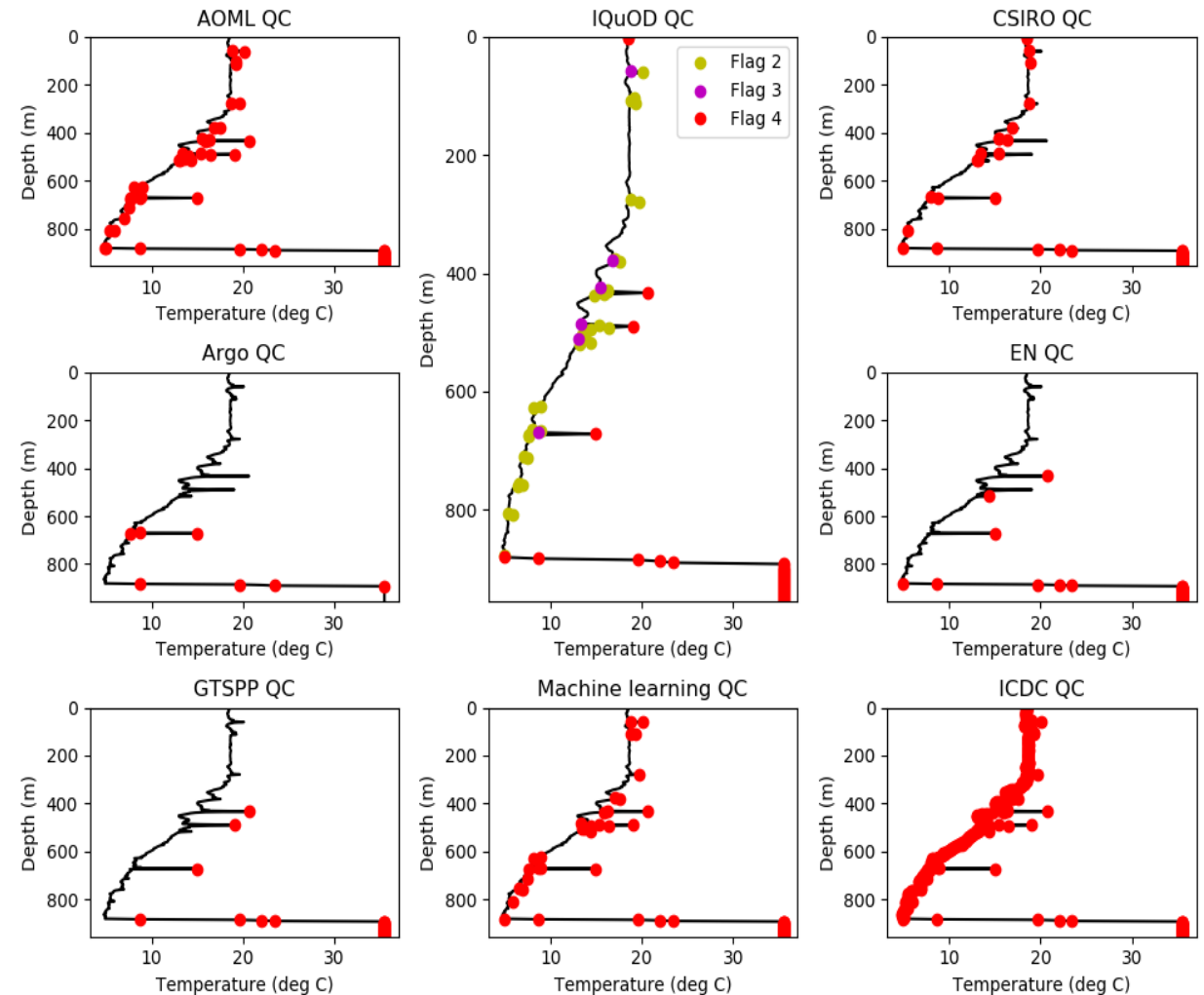
AQC tests performance



Our idea is to have three sets of QC tests. These will be used to set QC flags as follows:

- Flag 1 – good data.
- Flag 2 – rejected by our ‘High true positive rate’ QC set – this was generated to find the maximum amount of bad data that we can find, but with the drawback that there will be a decent amount of good data flagged also.
- Flag 3 – rejected by our ‘Compromise’ QC set – this tries to strike a balance between finding a lot of bad data without rejecting too much good data.
- Flag 4 – rejected by our ‘Low false positive rate’ QC set – emphasises reducing flagging of good data to a minimum but some bad data are missed.
- Difference applications may have different sensitivity to bad data and this flagging scheme will allow users to select the balance between TPR and FPR that they want.

Profile 060
North Pacific (x many spikes at many depths)



Benchmarking: 55 AQC tests

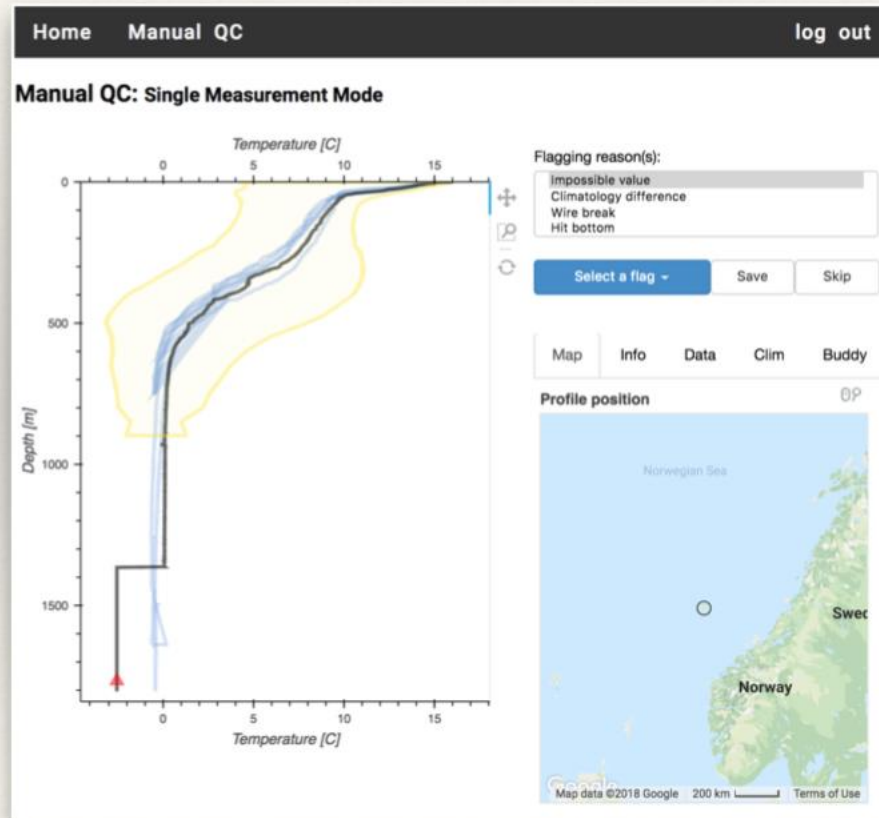
Future AQC work

Future aims would include:

- Use the QC sets in a future IQuOD data release.
- Identify why some profiles are not flagged by any test and develop methods for detecting those bad data.
- Work on optimising tests e.g. it looks like there may be benefit from optimising climatology checks.
- Use these metrics to monitor improvements in the performance of the machine learning algorithms achieved via the crowd source approach discussed at the workshop.

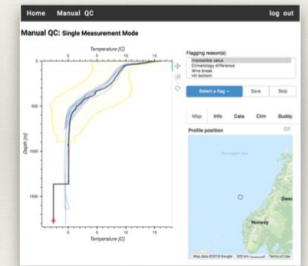
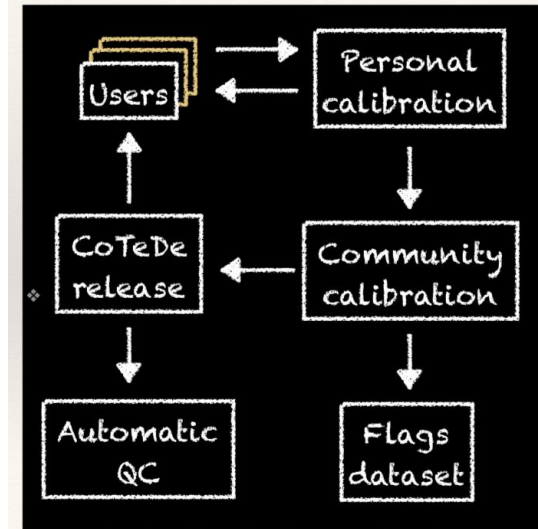
Expert QC user interface

<https://expertqc.castelao.net>



Community QC

<https://expertqc.castelao.net>



Products

- ❖ Close the cycle by returning the calibrated CoTeDe to the community
- ❖ Provide open access to the flagging dataset, so it can be used to calibrate other techniques

Expert QC

Future work

- ❖ Recruit, engage, and learn from QC experts (in progress);
- ❖ Employ advanced / alternative techniques for specific problems:
 - ❖ Wire break / Hit bottom for XBT (Rebecca Cowley- CSIRO);
 - ❖ Pattern identification (Ruth Musgrave - WHOI)
- ❖ Anonymize human flags and make them open access;
- ❖ Propose a procedure for (cross-)validation of experts (trainees);
- ❖ Training / educational: Teach new quality controllers to identify patterns;

n

International Quality-controlled Ocean Database

Repositories 7 People 3 Projects 0

Grow your team on GitHub
GitHub is home to over 28 million developers working together. Join them to grow your own development teams, manage permissions, and collaborate on projects.

[Sign up](#)

Find a repository... Type: All - Language: All -

IQuOD.github.io
International Quality-controlled Ocean Database
HTML 1 v3 Updated 5 days ago

AutoQC
A testing suite for automatic quality control checks of subsurface ocean temperature observations
Python 13 v8 MIT Updated 29 days ago

wodpy
A package to consume WOD format data.
Python 5 v3 MIT Updated on 1 Jul

Formats
Data format development
Updated on 18 Apr

Uncertainties
Uncertainty application to IQuOD temperature, depth and salinity values
Matlab Updated on 16 Apr

iMetadata
Coding to assign intelligent metadata
HTML Updated on 15 Jun 2017

machinelearn
Forked from oceanaustin/machinelearn
Machine Learning for XBT data
Python 2 v2 Updated on 6 Apr 2017

Top languages
Python HTML
Matlab

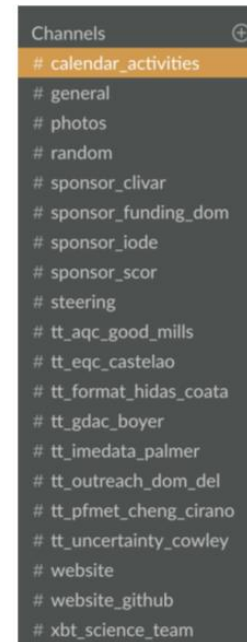
People 3 >
BillMills
Guilherme Castela
Marty Hidas

GitHub



Join by clicking at:

https://join.slack.com/t/iquod/shared_invite/enQtMzQ5NTFwNzAwNTMvLTBjZTNiMjY2E4ZDdiYThlZTk0MzQ3ZDZiZkZic1MzYxMmE2MikwM2NlNiI2ZTI5ZDg1NGVhZDNIYWE



www.iquod.org (being updated)

Workshop reports: <http://iquod.github.io/documents.html>

We would like to expand our IQuOD team & diversity

Task team leaders:

- **GDAC:** Tim Boyer
- **FORMAT:** Marty Hidas, Christine Coatanoan
- **Duplicates:** Ed King
- **I-metadata:** Matt Palmer, Toru Suzuki, Shoichi Kizu
- **Uncertainty:** Bec Cowley
- **AQC:** Simon Godd, Bill Mills
- **EQC:** Gui Castelao
- **Crowdsourcing:** Allison Macdonald, Steve Diggs
- **Performance metrics/User interface:** Lijing Cheng, Mauro Cirano
- **Funding/Outreach/Sponsors:** Catia Domingues, Janet Sprintall, Bec Cowley



Join us!

Email: iquod@outlook.com

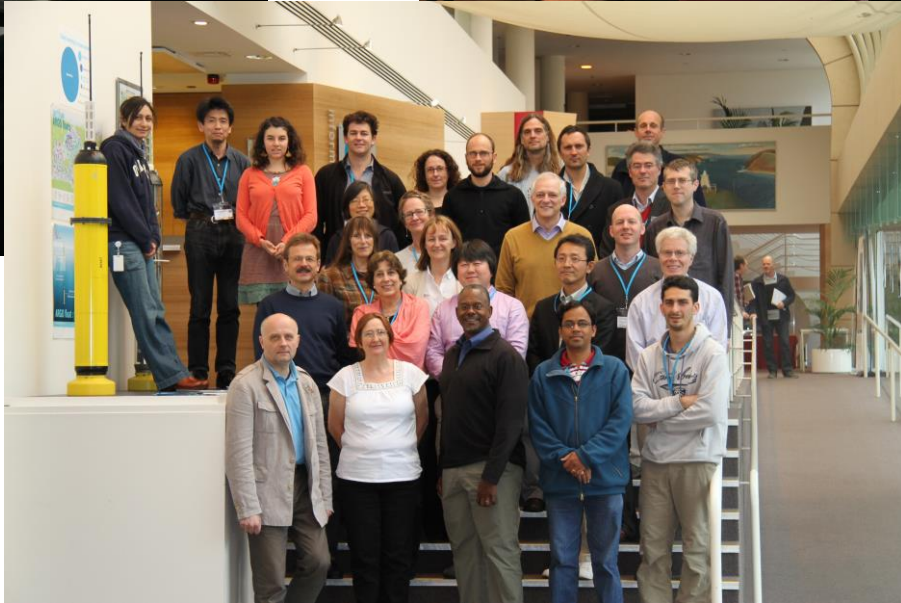


Many ways to contribute:

User/feedback, regional ocean expertise, include original data/metadata (via NCEI), AQC/EQC/Duplicates codes, Python programming, reference datasets, instrumentation and/or bias correction expertise, spread the word! etc.

Join us!





Thank you

