I Q u O D

SCOR WG 148 - IQuOD

International Quality-Controlled Ocean Database: Subsurface Temperature Profiles

To maximize the quality, consistency and completeness of our data

SCOR Annual Virtual Meeting,
26-28 October 2021

GSOP

SG-IQuOD

WG 148

https://scor-int.org/group/148/

www.iquod.org

7 ToRs & 7 Active task teams

ToR 6	• Global Data Assembly Center (GDAC) • Formats
ToR 5	• Uncertainty
ToR 1	• Inteligent Metadata
ToRs 2 & 3	• AutoQC | Duplicates
ToR 4	• ExpertQC | Machine Learning • Metrics
ToR 7	Share knowledge and transfer skills

IQuOD

www.iquod.org

Regular virtual meetings
No in person workshops

IQuOD

www.iquod.org

Publications 2020-2021

International Quality-controlled Ocean Database (IQuOD)

FOLLOW

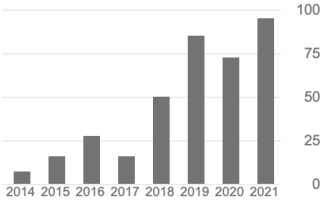
International Quality-controlled Ocean Database
Verified email at csiro.au - Homepage

oceanography ocean heat content ocean observations ocean data quality control

TITLE	CITED BY	YEAR
A machine learning approach to quality control oceanographic data GP Castelao Computers & Geosciences, 104803		2021
International Quality-controlled Ocean Database (IQuOD) v0. 1: the temperature uncertainty specification R Cowley, RE Killick, T Boyer, V Gouretski, F Reseghetti, S Kizu, ... Frontiers in Marine Science 8, 607	1	2021
A framework to quality control oceanographic data GP Castelao Journal of Open Source Software 5 (48), 2063	3	2020

Cited by

	All	Since 2016
Citations	383	350
h-index	9	9
i10-index	9	9



Co-authors

IQuOD Publications included in Ocean Best Practices

The IQuOD bibliography is maintained under a Google Scholar profile	[link]
IQuOD Ocean Best Practices Community	[link]
SCOR Ocean Best Practices Community	[link]

http://www.iquod.org/publications.html

Progress: ToR #1 | Intelligent metadata

To develop, implement and document algorithms for assignment of "intelligent" metadata – i.e. an informed guess as to likely values for missing information – for temperature profiles where crucial metadata is missing.

- Since Palmer et al. (2018), focus on machine learning techniques
  Python codes: https://github.com/MetOffice/XBTs_classification
- **ML techniques** tested and shown to outperform the existing algorithm
- **Paper in prep**. to be sumitted in 2021: "Improved infilling of missing metadata from eXpendable BathyThermographs using multiple machine learning methods"
- **Virtual meeting**: Results and future ideas discussed

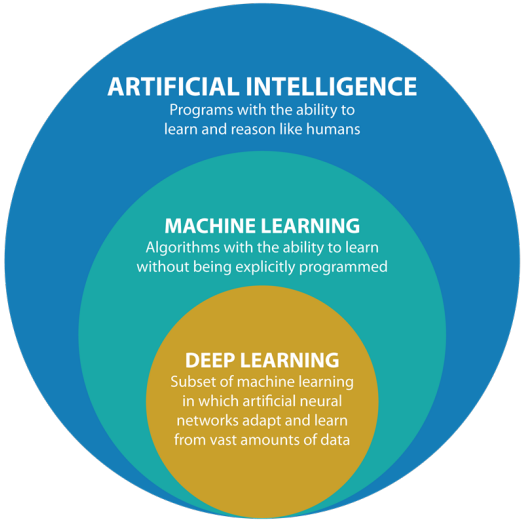
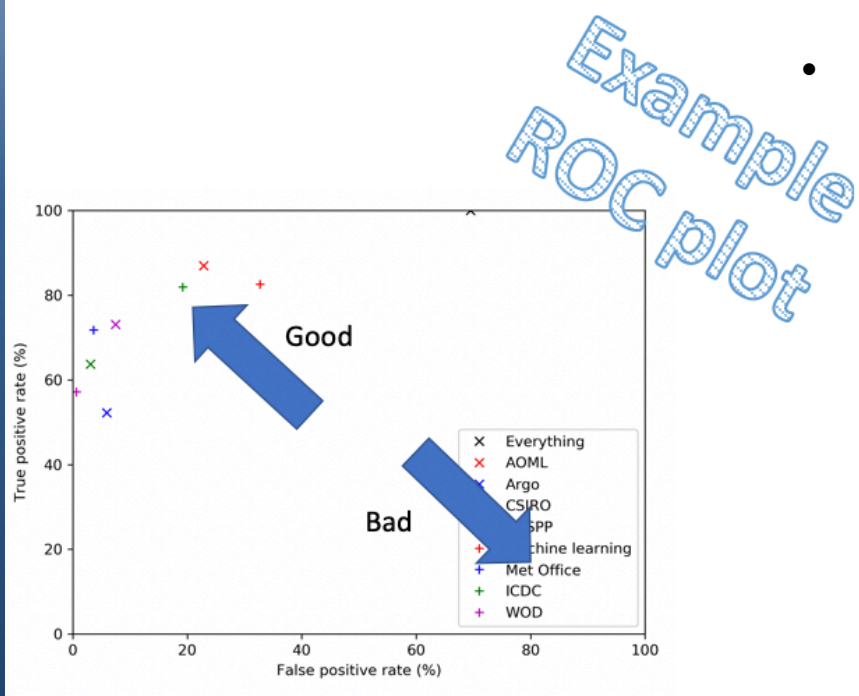


Image source: https://www.ufsm.br/pet/sistemas-de-informacao/2021/05/11/introducao-a-machine-learning/

Progress: ToRs #2 | AutoQC tests

To evaluate and document the most effective combination of automated quality control (AutoQC) procedures for temperature profile observations. International collaboration will be required for the design and coordination of benchmarking experiments using high-quality reference datasets.



- **Software suite** developed for benchmarking QC checks to find optimum set, using the highly QCed QuOTA dataset 2016 version with ~50K profiles (Gronnel and Wijffels, 2008).

Python codes: https://github.com/IQuOD/AutoQC (including Python reader for WOD native ASCII data)

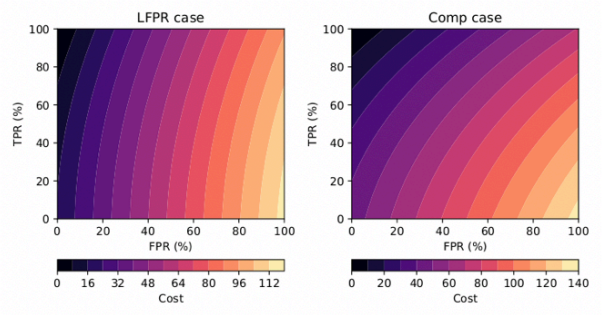
Progress: ToRs #3 | AutoQC optimum set

To establish and implement a set of optimal automated quality control procedures, by reaching international community consensus and using the knowledge gained in the benchmarking tests from ToR-2; to produce and publish a reference guide for best practices in automated quality control of ocean temperature profiles; and to develop and freely distribute an open-source quality control software toolkit to promote wide and rapid adoption of best practices by the oceanographic community.

Finding optimum test combinations

The LFPR and Comp cases are chosen by selecting at least one quality control check from each of the main types (range check, climatology check, etc.)

The choice of each QC check is determined by a tuneable cost function, which allows the operator to choose how much they wish to balance maximising high TPR over keeping FPR low



- Good et al. (2021) in prep.
- Open access GitHub codes

Progress: ToRs #4 | Expert QC ML

To examine and document the feasibility of machine learning and other novel computational methods for enhanced quality control, to potentially minimize labour costs associated with human expert quality-control procedures.

ELSEVIER

Computers & Geosciences
Volume 155, October 2021, 104803

Research paper

A machine learning approach to quality control oceanographic data

G.P. Castelão

Show more

Share Cite

https://doi.org/10.1016/j.cageo.2021.104803

Highlights

- Bad data can be identified from an anomalous behavior of its characteristics.
- A multivariate criterion overcomes traditional multiple univariate tests.
- Anomaly detection reduces the automatic quality control errors by at least 50%.
- Machine learning can optimize the expert quality control effort.
- CoTeDe is a Python package to apply the state of the art in quality control.

JOSS
The Journal of Open Source Software

A Framework to Quality Control Oceanographic Data

Guilherme P. Castelao[1]

1 Scripps Institution of Oceanography

DOI: 10.21105/joss.02063

Software
- Review
- Repository
- Archive

Editor: Kristen Thyng

Reviewers:
- @jessicaaustin
- @evanleeturner

Submitted: 26 December 2019
Published: 07 April 2020

Why GitHub? Team Enterprise Explore Marketplace Pricing Search Sign in Sign up

Security Insights

master 3 branches 11 tags Go to file Code

castelao Bump version: 0.23.7 → 0.23.8 3f11428 on 1 Jun 1,079 commits

.github/workflows	Fuzzylogic (#60)	10 months ago
cotede	Bump version: 0.23.7 → 0.23.8	5 months ago
docs	Notebooks on how to QC (#55)	6 months ago
joss	Moving WOD 2018 reference back to incollection	2 years ago
sampledata	Sample datasets (#54)	10 months ago
tests	feat: Expects attrs instead of attributes (#64)	5 months ago
.gitattributes	Setting up LFS: *.gz|nc|cdf	6 years ago
.gitignore	Git now ignores .swp (vim swap files)	6 years ago
.travis.yml	Adding 3.8 into travis' checks.	15 months ago
.zenodo.json	doc: Extending references	6 months ago
AUTHORS.rst	fix: Explicit call to qctests.Morello2014 (#61)	9 months ago
CITATION.cff	Bump version: 0.23.7 → 0.23.8	5 months ago
CONTRIBUTING.rst	BUGFIX #40, do not recommend flake8 anymore.	2 years ago
HISTORY.rst	Updating and cleanning documentation	2 years ago
LICENSE.rst	Adjusting license details.	6 years ago
MANIFEST.in	Moving installing requirements to requirements.txt	6 years ago
Makefile	Release new versions to pypi using twine.	5 years ago
README.rst	doc: Extending references	6 months ago
environment.yml	Notebooks on how to QC (#55)	6 months ago
readthedocs.yml	Using conda to build documentation (readthedo...	5 years ago
requirements.txt	Flexible WOA comparison (#57)	10 months ago
setup.cfg	Bump version: 0.23.7 → 0.23.8	5 months ago
setup.py	Bump version: 0.23.7 → 0.23.8	5 months ago
test-requirements.txt	Fuzzylogic (#60)	10 months ago
tox.ini	perf: Updating OceansDB and testing it	13 months ago

README.rst

CoTeDe

About
Quality Control of Oceanographic Data
Readme
BSD-3-Clause License
Cite this repository

Releases 5
JOSS paper Latest on 31 Mar 2020
+ 4 releases

Packages
No packages published

Used by 1
@BjerknesClimateData...

Contributors 4
castelao Guilherme Cast...
BillMills Bill Mills
kthyng Kristen Thyng
s-good Simon Good

Languages
Python 98.3%
Other 1.7%

Progress: ToRs #5 | Uncertainty temperature

To develop, implement and document internationally agreed best practice methods for assignment of uncertainty estimates to each temperature observation.

OceanBestPractices
Repository of Ocean Community Practices in Ocean Research, Observation and Data/Information Management

OceanBestPractices Home / ⇒ IQuOD:International Quality-controlled Ocean Database Project

⇒ IQuOD:International Quality-controlled Ocean Database Project

BROWSE BY

By Issue Date | Authors | Titles | Subjects

Search within this community and its collections:

Go

The IQuOD (International Quality-controlled Ocean Database) effort is being organized by the oceanographic community, and includes experts in data quality and management, climate modellers 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 along with (intelligent) metadata and assigned uncertainties for use in ocean climate research applications. This goal will be achieved by developing and implementing an internationally-agreed framework.

http://www.iquod.org/

Collections in this community

IQuOD Community Practices [2]

Recent Submissions

International Quality-Controlled Ocean Database (IQuOD) v0.1: The Temperature Uncertainty Specification.

Cowley, Rebecca; Killick, Rachel E.; Boyer, Tim; Reseghetti, Franco; Kizu, Shoichi; Palmer, Matthew D.; Cheng, Lijing; Storto, Andrea; Le Menn, Marc; Simoncelli, Simona; Macdonald, Alison M.; Domingues, Catia M. (2021)

Ocean temperature observations are crucial for a host of climate research and forecasting activities, such as climate monitoring, ocean reanalysis and state estimation, seasonal-to-decadal forecasts, and ocean forecasting. ...

frontiers in Marine Science | Ocean Observation

SECTION ABOUT ARTICLES RESEARCH TOPICS FOR AUTHORS EDITORIAL BOARD ARTICLE ALERTS

‹ Articles

THIS ARTICLE IS PART OF THE RESEARCH TOPIC
Best Practices in Ocean Observing View all 34 Articles ›

EDITED BY
Jay S. Pearlman
Institute of Electrical and Electronics Engineers (France), France

REVIEWED BY
Mark Bushnell
National Ocean Service (NOAA), United States

Gabriele Pieri
Istituto di Scienza e Tecnologie dell'Informazione 'Alessandro Faedo' (ISTI), Italy

The editor and reviewers' affiliations are the latest provided on their Loop research profiles and may not reflect their situation at the time of review.

TABLE OF CONTENTS

ORIGINAL RESEARCH article
Front. Mar. Sci., 11 June 2021 | https://doi.org/10.3389/fmars.2021.689695

International Quality-Controlled Ocean Database (IQuOD) v0.1: The Temperature Uncertainty Specification

Rebecca Cowley[1,2*], Rachel E. Killick[3], Tim Boyer[4], Viktor Gouretski[5], Franco Reseghetti[6], Shoichi Kizu[7], Matthew D. Palmer[3,8], Lijing Cheng[9], Andrea Storto[9], Marc Le Menn[10], Simona Simoncelli[11], Alison M. Macdonald[12] and Catia M. Domingues[13,14]

[1]Oceans and Atmosphere, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Hobart, TAS, Australia
[2]Centre for Southern Hemisphere Oceans Research, Hobart, TAS, Australia
[3]Met Office Hadley Centre, Exeter, United Kingdom
[4]NOAA National Centers for Environmental Information, Silver Spring, MD, United States
[5]International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[6]Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Santa Teresa Research Centre, Pozzuolo di Lerici, Italy
[7]Physical Oceanography Laboratory, Department of Geophysics, Tohoku University, Sendai, Japan
[8]University of Bristol, Bristol, United Kingdom
[9]Institute of Marine Sciences, National Research Council, Rome, Italy
[10]Service Hydrographique et Océanographique de la Marine, Brest, France
[11]Istituto Nazionale di Geofisica e Vulcanologia (INGV), Bologna, Italy
[12]Physical Oceanography Department, Woods Hole Oceanographic Institution, Woods Hole, MA, United States
[13]National Oceanography Centre, Southampton, United Kingdom
[14]Australian Research Council (ARC) Centre of Excellence for Climate Extremes, University of Tasmania, Hobart, TAS, Australia

Ocean temperature observations are crucial for a host of climate research and forecasting activities, such as climate monitoring, ocean reanalysis and state estimation, seasonal-to-decadal forecasts, and ocean forecasting. For all of these applications, it is crucial to understand the uncertainty attached to each of the observations, accounting for changes in instrument technology and observing practices over time. Here, we describe the rationale behind the uncertainty specification provided for all in situ ocean temperature observations in the International Quality-controlled Ocean Database (IQuOD) v0.1, a value-added data product served alongside the World Ocean Database (WOD). We collected information from manufacturer specifications and other publications, providing the end user with uncertainty estimates based mainly on instrument type, along with extant auxiliary information such as calibration and collection method. The provision of a consistent set of observation uncertainties will provide a more complete understanding of historical ocean observations used to examine the changing environment. Moving forward, IQuOD will continue to work with the ocean observation, data assimilation and ocean climate communities to further refine uncertainty quantification. We encourage submissions of metadata and information about historical practices to the IQuOD project and WOD.

Progress: ToRs #6 | GDAC

To freely disseminate (interim) versions of the IQuOD global temperature profile database (and added-value products) as it evolves over the next 3 years, in user-friendly file formats.

International Quality Controlled Ocean Database (IQuOD) version 0.1 - aggregated and community quality controlled ocean profile data 1772-2018 (NCEI Accession 0170893)

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 along with (intelligent) metadata and assigned uncertainties for use in ocean climate research applications. This goal will be achieved by developing and implementing an internationally agreed framework. IQuOD v0.1 is a preliminary data set which includes uncertainties on each temperature measurement and intelligent metadata for identifying critical missing information.

Show less

Dataset Citation

Dataset Identifiers

ISO 19115-2 Metadata

Access | Time & Location | Documentation | Description | Credit | Keywords | Constraints | Lineage

Download Data	HTTPS *(download)* Navigate directly to the URL for data access and direct download. FTP *(download)* These data are available through the File Transfer Protocol (FTP). FTP is no longer supported by most internet browsers. You may copy and paste the FTP link to the data into an FTP client (e.g., FileZilla or WinSCP). THREDDS *(download)* 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/. HTTPS *(download)* Navigate directly to the URL for data access and direct download. FTP *(download)* These data are available through the File Transfer Protocol (FTP). FTP is no longer supported by most internet browsers. You may copy and paste the FTP link to the data into an FTP client (e.g., FileZilla or WinSCP).
Distribution Formats	• netCDF (Version: netCDF-4) ◦ File Specification: IQuOD Multi-cast file
Ordering Instructions	Contact NCEI for other distribution options and instructions.
Distributor	*NOAA National Centers for Environmental Information* NCEI.Info@noaa.gov
Dataset Point of Contact	*NOAA National Centers for Environmental Information* ncei.info@noaa.gov

Last Modified: 2021-10-19T21:32:49Z

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

National Centers for Environmental Information
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

FOLLOW US
@NOAANCEI
@NOAAData
Ocean & Geophysics
Weather & Climate

CONTACT US
Email: ncei.info@noaa.gov
Phone: (828) 271-4800

https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:0170893

Progress: ToRs #7 | Share knowledge & Skills

To share knowledge and transfer skills in instrumentation, regional oceanography, quality control procedures and data stewardship with international scientists in both developed and developing nations.

- Virtual meetings
  - Member discussions
  - Invited expert presentations and/or contributions
- Code & data & documentation available online
  - IQuOD.org
  - GitHub (IQuOD, CoTeDe, etc)
  - NCEI/NOAA
  - Best ocean community practices repositories
  - Google scholar
- Peer-reviewed (preferably open-access) publications

Ongoing planned activities

- Continue **regular virtual meetings** (online calendar list available for 2021/22). In person workshop maybe considered depending on pandemic situation in 2022/23. Talks at international conferences (e.g. OSM 2022).

- Submission/publication of **AutoQC checks paper**

- Release of **new IQuOD dataset version based on optimum AutoQC set** applied to latest WOD version

- **Improvement to Auto & Expert QCs** by developing and implementing further training data & machine learning techniques. Cloud computing activities being sponsored by IOC/IODE.

- Metrics task team and end-users: **Uptake and performance of IQuOD dataset version releases**

- **Seeking further funding** (e.g. NERC UK/NFS USA). **Seeking letters of support**

- **Seeking endorsement** from UN Decade of Ocean Science for Sustainable Development.

IQuOD

Many thanks

to the SCOR committee

for their support