IQuOD Annual SCOR Working Group Report to SCOR

1. Name of group

Working Group 148: International Quality controlled Ocean Database (IQuOD)

2. Activities since previous report to SCOR (e.g., virtual or in-person meetings, email discussions, special sessions). Limit 1000 words

Due to the pandemic, there have been no in person meetings during the last year. Instead, virtual meetings have been held covering topics including funding and interpolation of ocean profiles.

Work has progressed on a range of activities. In particular, a paper has been published on assignment of uncertainties to ocean temperature profiles, and work is ongoing on extending a draft paper on benchmarking of automatic quality control checks with publication anticipated during the coming year. In addition, a community for IQuOD has been set up in the Ocean Best Practices Repository (https://repository.oceanbestpractices.org/handle/11329/1590).

3. Documents published since previous report to SCOR (e.g., peer-reviewed journal articles, reports, Web pages) and should be limited to publications that resulted directly from WG activities and which acknowledge SCOR support

Cowley, R., Killick R. E., Boyer T., Gouretski V., Reseghetti F., Kizu S., Palmer M. D., Cheng L., Storto A., Le Menn M., Simoncelli S., Macdonald A. M., Domingues C. M. (2021) International Quality-Controlled Ocean Database (IQuOD) v0.1: The Temperature Uncertainty Specification, Frontiers in Marine Science, 8, 607, doi: 10.3389/fmars.2021.689695.

4. Progress toward achieving group's terms of reference. List each term of reference separately and describe progress on each one. Limit 1000 words

1. 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.
   - Previously, IQuOD defined an algorithm to assign intelligent metadata for XBTs (Palmer et al., 2018; https://doi.org/10.1175/JTECH-D-17-0129.1). Since, IQuOD has leveraged resources at the Met Office to increasingly focus on machine learning techniques to improve upon the technique, with code available on GitHub (https://github.com/MetOffice/XBTs_classification). A number of different machine learning techniques have been tested and shown to outperform the existing algorithm. A paper ('Improved infilling of missing metadata from eXpendable BathyThermographs using multiple machine learning methods') is in preparation to describe the work, for submission in 2021. An IQuOD virtual meeting discussed the

results and a number of proposals for future work were discussed.

2. 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.
   - A software suite has been developed (https://github.com/IQuOD/AutoQC) that performs benchmarking of quality control checks using datasets identified by the IQuOD team and finds optimum sets of checks. This has previously been run to generate results and a draft paper written. Over the last year the work was extended to include more reference data and extra quality control checks. The software will be rerun imminently and the results used to update the draft paper. It is therefore anticipated that the paper will be published in the coming year.

3. 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 (above); 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.
   - The software described in ToR-2 is open source and is published under an open license (the MIT license). It is planned to publish the paper describing the software and the results from running it as open access.

4. To examine and document the feasibility of machine learning and other novel computational methods for enhanced quality control, to potentially minimize labor costs associated with human expert quality-control procedures.
   - Funding has been obtained to provide IQuOD with a cloud computing account to support this activity in the future.
   - A paper (Castelao, 2020; https://doi.org/10.21105/joss.02063) on the framework used to implement and aggregate different QC methods.
   - A paper (Castelao, 2021; https://doi.org/10.1016/j.cageo.2021.104803) on the machine learning technique used to guide and optimize the human expert effort.

5. To develop, implement and document internationally agreed best practice methods for assignment of uncertainty estimates to each temperature observation.
   - A paper (Cowley et al., 2021; https://doi.org/10.3389/fmars.2021.689695) has recently been published, which defines uncertainty estimates for temperature data.
   - A community has been set up in the Ocean Best Practices Repository for IQuOD (https://repository.oceanbestpractices.org/handle/11329/1590).

6. 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.
   - IQuOD v0.1 data are freely available from the US NCEI World Ocean Database website (https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:0170893) in the widely used netCDF format.

7. 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.
   - The paper on uncertainty estimates (Cowley et al., 2021; https://doi.org/10.3389/fmars.2021.689695) has been published as an open-access publication, allowing access to everyone without cost. It has also been submitted to Ocean Best Practices.

5. WG activities planned for the coming year. Limit 500 words

A primary focus of IQuOD in the next year is the publication of a paper describing benchmarking of automatic quality control checks for temperature data. The paper will also include recommendations for optimum sets of quality control checks that will be applied to the World Ocean Database (WOD) to generate a new version of the IQuOD dataset. A draft of the paper has already been completed, but will be updated to use more data and include more quality control checks before submission.

A second focus will be on developing training data and techniques for machine learning to improve quality control of data further. Funding has been obtained for cloud computing to support this activity.

It is anticipated that IQuOD meetings will continue to be online over the next year. Short meetings on different topics are expected to be held regularly and there will be discussions on whether a longer workshop is viable online given the disparate time zones that the IQuOD team live in.

6. Is the group having difficulties expected in achieving terms of reference or meeting original time schedule? If so, why, and what is being done to address the difficulties Limit 200 words

The pandemic has disrupted activities, for example by preventing meetings and limiting time available for working on the project. In the former case, online meetings have been instigated in order to partially compensate. A full workshop has not been held online, but the smaller meetings have assisted with continuing to move things forward.

7. Any special comments or requests to SCOR. Limit 100 words.

Additional information can be submitted and will be included in the background book for the SCOR meeting at the discretion of the SCOR Executive Committee Reporter for the WG and the SCOR Secretariat.