WG 148 - Annual SCOR Working Group Report

Summary:

IQuOD is the International Quality Controlled Ocean Database. It aims to maximize the quality, consistency and completeness of the long-term global subsurface ocean temperature profile record by bringing together scientists and software engineers from across the world to collaborate and reduce duplication of effort. It disseminates the results of this work through papers, presentations and provision of a dataset to users.

Part of the IQuOD project is focussed on improving automatic quality control of ocean temperature profiles. There are many quality control checks in use around the world but the effectiveness of these is not known. In the last year, a draft paper describing IQuOD's work on benchmarking the performance of automatic quality control checks has been revised and is currently undergoing additional revisions following further comments from co-authors. In addition, activity in an IQuOD task team focussed on detecting duplicate profiles has increased during the last year. This collaboration has begun by understanding methods that are in use in different institutions. Regular virtual meetings have been held to foster these, and other, IQuOD activities.

Next year, the primary focus of IQuOD will be the publication of the paper describing benchmarking of automatic quality control checks. Once published, its recommendations for optimum sets of quality control checks will be applied to the World Ocean Database (WOD) to generate a new version of the IQuOD dataset. It is expected that other work will include developing training data and techniques for machine learning to improve quality control of data further, and continuation of the collaboration on detection of duplicate profiles.

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

The paper describing work on benchmarking automatic quality control checks has been revised and is currently undergoing additional revisions following further comments from co-authors. In the past year, activity in the task team focussed on detecting duplicate profiles has increased. This collaboration has begun by understanding methods that are in use in different institutions.

Due to the pandemic, there have been no in person meetings during the last year. Instead, virtual meetings have been held.

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

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.
    - Achieved: a paper was previously published on this by IQuOD and the IQuOD dataset already includes the results of applying its intelligent metadata for expendable bathythermograph (XBT) data.

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.
    - Revisions to a draft paper about this work have been made. Following further revision, 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.
    - A cloud computing account to support this activity continues to be funded; ongoing work includes transferring code to this cloud facility.

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

- Achieved: a paper was previously published on this by IQuOD and the IQuOD dataset already includes those uncertainty estimates.

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.
   - IQuOD virtual meetings are now being held on a monthly basis, in three alternating time slots across the day to increase accessibility of the meetings for scientists across the globe.

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

The primary focus of IQuOD continues to be the publication of a paper describing benchmarking of automatic quality control checks for temperature data. Once published, its recommendations for optimum sets of quality control checks will be applied to the World Ocean Database (WOD) to generate a new version of the IQuOD dataset.

A second focus will be on developing training data and techniques for machine learning to improve quality control of data further.

Thirdly, methods for detection of possible duplicate profiles will continue to be developed and working practices established to determine if detected profiles are true duplicates.

It is anticipated that an in-person IQuOD meeting will be held next year. In addition, short virtual meetings on different topics are expected to be held regularly.

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.

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.