

General comments:

The authors have much improved the manuscript since the first review. However, there are still some issues that both myself and the other reviewer highlighted that have not been fixed. These issues are associated with the data and metadata via ERDDAP and should be fixed before publication.

Despite the author's wish to follow FAIR principles, the data is not accessible in its current format as the entire dataset cannot be downloaded.

The attempt to make the metadata available using a separate set of url links in the main data file is not satisfactory. The effort that the authors have put into recovery of this information should lead to the metadata being included in the file as variables, not as a secondary data file only accessible via url coding.

One final suggestion is that the authors make the automatic QC code used available publicly, perhaps via a git repository.

Answer:

We thank the reviewer for the detailed review which allows us to further improve the dataset and make it fit for use. This is crucial for us that decided to adopt ERDDAP server, a FAIR-compliant data access service (O'Brien and Delaney, 2024), in line with the GOOS (Global Ocean Observing System) Observations Coordination Group (<https://goosocean.org/who-we-are/observations-coordination-group/>) strategy. In fact, according to Lange et al. (2023), ERDDAP "(i) supports dozens of popular formats; (ii) provides standards-based metadata and data services and formats; (iii) supports federated access of distributed ERDDAP data services; (iv) supports both human and machine interactions; (v) supports sub-setting of large datasets; (vi) provides improved discovery of datasets through commercial search engines; and (vii) provides support for archival of datasets". (Text added to Section 7)

We are sorry that the data was not accessible, we have tried several times since your feedback and we had no problems in downloading the data. We hypothesize an incidental server load and traffic.

We are also sorry that the availability of all existing metadata as a URL link is not satisfactory for the reviewer. The decision to provide metadata in a separate metadata URL has been taken in collaboration with the [EMODnet Physics](#) data managers within the framework of [EMODnet Data Ingestion](#) project (Novellino et al., 2024). In fact, the REP XBT dataset is also accessible through EMODnet Physics data portal, thanks to the ERDDAP machine to machine interoperability. Including all metadata as variables in the file, would make the ERDDAP conversion to the dozens output formats less efficient due to the many metadata strings to manage.

To meet the reviewer request and facilitate the data reusability we prepared a [Jupyter Notebook](#) in Python that allows recombining all data and metadata in NetCDF files, one per XBT profile. The notebook is available on a GitHub repository and published on Zenodo, in compliance with Open Science principles.

The publication of the automatic QC code was out of our scope when reprocessing and analyzing the XBT data, so we did not work to prepare a FAIR software (Baker et al., 2022) but to prepare a FAIR dataset. We believe that sharing a code which is not fit-for-use at this stage of the review

process would not be beneficial for both producers and users. We would like, if accepted the dataset paper, to work on optimizing the code and making it FAIR. We think that publishing it in an open software journal could help to improve it by going through a peer review.

Barker, M., Chue Hong, N.P., Katz, D.S. et al. *Introducing the FAIR Principles for research software*. *Sci Data* 9, 622 (2022). <https://doi.org/10.1038/s41597-022-01710-x>.

Lange N, Tanhua T, Pfeil B, Bange HW, Lauvset SK, Grégoire M, Bakker DCE, Jones SD, Fiedler B, O'Brien KM and Körtzinger A (2023) A status assessment of selected data synthesis products for ocean biogeochemistry. *Front. Mar. Sci.* 10:1078908. doi: 10.3389/fmars.2023.1078908

Novellino, A., Pizziol, V., Dapueto, G., Misurale, F., Scotto, B. M., Bordoni, R., Gorringer, P., Schaap, D., & Iona, A. (2024). *EMODnet Ingestion and the operational data exchange examples and hot topics*. *Miscellanea INGV*, 80, 364–366. <https://doi.org/10.13127/MISC/80/140>

O'Brien, K., & Delaney, C. (2024). *A review of ERDDAP the established best practice in sharing gridded and tabular data from the Earth Sciences community*. *Miscellanea INGV*, 80, 231–232. <https://doi.org/10.13127/MISC/80/87>

Manuscript specific comments:

Line 210-217: The term ‘accuracy’ is used for the XCTDs. As these values are translated to ‘uncertainty’ in the data files, you need to explain that you have used the ‘accuracy’ values as type B uncertainty, as done in the Cowley et al uncertainty paper.

Answer: Text has been added: *“These accuracy values can be considered Type B uncertainties, as in Cowley et al. (2021), and they are included in the REP dataset metadata information.”*

Line 375–379: Position on land check. I asked about this in the previous review and your response was: *“There are no profiles on land in the REP dataset, since the operators checked both the position and the launch time before the data transmission to the ENEA-STE. Since we did not encounter specific issues with date/time we did not implement additional checks.”* I suggest you include this information in the paper to give the reader reassurance that there are no data flagged as bad due to position on land and that position errors were already fixed prior to this test.

Answer: We thank the reviewer for this suggestion, we added a phrase in the manuscript.

Figure10: please add more description to the figure caption so the reader doesn’t have to refer back to the text. Or highlight individual issues with circles/shapes on the figure.

Answer: We thank the reviewer for the suggestion, some text has been added to the caption and some shapes have been added on the plots. Some text has been eliminated at page 26 too.

Data comments:

- I downloaded a sub-setted netcdf file. It still has global attributes that should be removed from the global attributes section, and many need review. In particular, the green highlighted attributes belong to individual profiles and should be translated to variables. The yellow highlighted attributes are repeats of other attributes and include spelling mistakes.

Answer: The metadata reported by the reviewer have been subdivided into blocks to manage specific answers.

Spelling mistakes have been corrected.

```
// global attributes:
:sea = "Mediterranean Sea";
:bathymetric_information = "102.m"; :cdm_data_type = "Profile";
:cdm_profile_variables = "latitude,longitude,time,profile_id";
:citation = "Reseghetti F., Fratianni C., Simoncelli S. (2024). Reprocessed of XBT dataset in the Ligurian and Tyrrhenian seas (1999-2019) (Version 2) [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/rep\_xbt\_1999\_2019.2";
:contributor_name = "Franco Reseghetti, Claudia Fratianni, Simona Simoncelli, Antonio Guarnieri";
:contributor_role = "Data Collector, Data Curator/Manager, Researcher/Supervisor, Project Leader";
:contributor_url = "orcid.org/0000-0002-7569-8541, orcid.org/0000-0002-4983-4255, orcid.org/0000-0001-8162-6571";
:conventions = "SeaDataNet_1.0_CF-1.6_COARDS_ACDD-1.3";
:creator_name = "Istituto Nazionale di Geofisica e Vulcanologia, Sede di Bologna (INGV)";
:creator_url = "https://www.ingv.it/";
:data_mode = "D";
:depth_uncertainty = "depth <= 230m: 4.6m; depth >= 230m: 2% (Table 2 from Cowley R et al., 2021 https://doi.org/10.3389/fmars.2021.689695)";
:doi = "https://doi.org/10.13127/rep_xbt_1999_2019.2";
:feature_type = "Profile";
:history = "2024-07-08T02:56:09Z (local file) # "2024-07-08T02:56:09Z https://oceanic.bo.ingv.it/ndap/tabledap/REP\_XBT\_1999\_2019\_v2.nc CALIB%2CDEPTH%2CDEPTH_COR_FLAGS_QC%2CDEPTH_COR_INT%2CDEPTH_COR_INT_SEADATANET_QC%2CDEPTH_FLAGS_QC%2CDEPTH_INT%2CDEPTH_INT_SEADATANET_QC%2CDEPTH_TEST_QC%2Clatitude%2Clongitude%2CPOSITION_SEADATANET_QC%2CSDN_BOT_DEPTH%2CSDN_CRUISE%2CSDN_EDMO_CODE%2CTEMPET01%2CTEMPET01_COR_FLAGS_QC%2CTEMPET01_COR_INT%2CTEMPET01_COR_INT_SEADATANET_QC%2CTEMPET01_FLAGS_QC%2CTEMPET01_FLAGS_QC%2CTEMPET01_INT%2CTEMPET01_INT_SEADATANET_QC%2CTEMPET01_TEST_QC%2Ctime%2CTIME_SEADATANET_QC%2Curl_metadata%2Ctime%2C-2019-09-28T09%3A21%3A13Z";
:infoUrl = "https://www.ingv.it/it/http://www.ingv.it/";
:institution = "ENEA Centro Ricerche Adriatico Mareno - La Spezia, Istituto Nazionale di Geofisica e Vulcanologia, Sede di Bologna (INGV)";
:keywords = "1999-2019, ambiente, applied, applying, baker, bathymetric, bathymetry, bathythermograph, below, bolegna, CALIB, calibrated, calibration, cdi, centro, cheng2014, code, control, cor, corrected, cruise, cruise_id, data, dataset, depth, DEPTH_COR, DEPTH_COR_FLAGS_QC, DEPTH_COR_INT, DEPTH_COR_INT_SEADATANET_QC, DEPTH_FLAGS_QC, DEPTH_INT, DEPTH_INT_SEADATANET_QC, DEPTH_TEST_QC, directory, each, earth, Earth Science > Oceans > Ocean Temperature > Water Temperature, earth science-oceans-ocean temperature-ocean temperature profiles, enea, european, expendible, flag, flag, floor, geofisica, grouping, ingv, international, istituto, label, latitude, levels, liguria, longitude, MACMAP, mapping, marine, marino, measurement, method, nazionale, ocean, oceanic, organizations, original, partner, POSITION_SEADATANET_QC, profile, profile_id, quality, rep_xbt_1999_2019_v2, reprocessed, revolution, resulting, ricerca, science, SDN, BOT, DEPTH, SDN, CRUISE, SDN, EDMO_CODE, sea, sea, floor, depth, below, sea, surface, sea, water, temperature, seadatanet, seafloor, seas, seawater, side, site, spezia, standard, status, status, flag, surface, temperature, tempet01, TEMPET01_COR, TEMPET01_COR_FLAGS_QC, TEMPET01_COR_INT, TEMPET01_COR_INT_SEADATANET_QC, TEMPET01_FLAGS_QC, TEMPET01_INT, TEMPET01_INT_SEADATANET_QC, TEMPET01_TEST_QC, test, tyrrhenian, time, TIME_SEADATANET_QC, topography, using, vertical, vulcanologia, water, xbt";
:keywords_vocabulary = "ICMD Science keywords";
:netcdf_version = "netCDF classic model";
:pi_name = "Franco Reseghetti";
:sourceId = "local files";
:standard_name_vocabulary = "CF Standard Name Table v70";
:subsetVariables = "CALIB, latitude, longitude, POSITION_SEADATANET_QC, SDN, BOT_DEPTH, SDN_CRUISE, SDN_EDMO_CODE, time, TIME_SEADATANET_QC, cruise_id, profile_id";
:summary = "The Rep_XBT_1999_2019 (V2) dataset comes from an improved and fully automatic quality control procedure applied to all available raw data and metadata from Expendable Bathythermograph \(XBT\) probes sampled and managed by ENEA S.Teresa Centre since September 1999 in the Ligurian and Tyrrhenian Seas. The reprocessed dataset contains a full metadata description obtained from the cruise reports according to the most recent community standards and formats. The second version of the dataset has been released as a final result of the revised manuscript";
:TEMPET01_uncertainty = "XBT = 0.10 deg C, MTD = 0.02 deg C (Table 2 from Cowley R et al., 2021 https://doi.org/10.3389/fmars.2021.689695)";
:title = "Reprocessed XBT dataset (V2) in the Ligurian and Tyrrhenian Seas (1999-2019)";
:license = "Attribution (CC BY 4.0)";
```

The below attributes, referring to each single profile, have been removed from global attributes and included in the url_metadata, which is a variable URL linking to the metadata information. Data and metadata of each profile can be easily associated through the *profile_id* and *cruise_id* fields.

```
:fall_rate_equation_Coeff_1 = "6.301m s^-1";
:fall_rate_equation_Coeff_2 = "0.00216m s^-2";
:family_code = "XB";
:ices_platform_code = "48AA";
:IMO_number = "8642751";
:launching_height = "05m";
:last_good_depth_according_to_operator = "92.78m";
:max_acquisition_depth = "300m";
:max_recorded_depth = "292m";
:platform_code = "IGMA";
:platform_name = "Ammiraglio Magnaghi";
:probe_manufacturer = "Lockheed Martin Sippican";
:probe_serial_number = "345751";
:probe_type = "T-10";
:qc_indicator = "excellent (all important QC done)";
:recorder_sampling_frequency = "10Hz";
:recorder_serial_number = "239";
:recorder_types = "72";
:ship_speed = "06kn";
:source_platform_category_code = "31";
:wmo_inst_type = "061";
:wmo_platform_code = "IGMA";
```

We decided to remove this information from global attributes since it is referring to internal identification of the profile:

```
:id = "T0_M1947";
```

The metadata attributes below have been removed since they are repeats of geospatial attributes.

```
:last_latitude_observation = "44.2966"
:last_longitude_observation = "9.1445"
```

The global metadata attributes below, refer to the entire dataset or to a portion of the selected dataset. The green ones are kept since they are part of the recommended Attribute Convention for Data Discovery 1-3 (ACDD). The geospatial yellow attributes are inserted by default in ERDDAP global metadata attributes to visualize the data via Google Earth, thus they cannot be removed.

```
:geospatial_lat_max = 44.2966 ;
:geospatial_lat_min = 43.8832 ;
:geospatial_lon_max = 9.1445 ;
:geospatial_lon_min = 8.8679 ;
```

```
:geospatial_vertical_max = 550.59f ;
:geospatial_vertical_min = 0.f ;
:time_coverage_end = ""2019-09-28T00:21:13Z"" ;
:time_coverage_start = ""2019-09-27T20:00:22Z"" ;
:Easternmost_Easting = 9.1445 ;
:Northernmost_Northing = 44.2966 ;
:Southernmost_Northing = 43.8832 ;
:Westernmost_Easting = 8.8679 ;
```

This is a ACDD global attribute referring to Publication information. It is not required but we decided to include it since it gives information about the dataset update. In our case it is “void” because the dataset is not updated on a schedule.

```
:update_interval = ""void"" ;
```

- Thanks for including uncertainties. I suggest 'depth_uncertainty' should be moved to an attribute in the Depth* variables or calculated out and added as its own variable. 'TEMPET01_uncertainty' should also be an attribute in the TEMPET01 variable and be specific to the instrument used (XBT or CTD). The reference can still be included as a global attribute.

Answer: As suggested by the reviewer, the uncertainty info has been moved to depth and TEMPET01 variables.

- The url_metadata link should be replaced with the actual data. To use the links, more coding is required (thanks for the example in C.6). Each link goes to a single line of information on a html page that could just as easily be included as variables in the datafile. I strongly suggest you put the information in these links into variables, especially since one of the main purposes of the paper is to rescue metadata and ensure it is attached to the data (see your own introduction), as well as make it Accessible.

Answer: The url_metadata is a URL variable that we provide in agreement with the EMODnet Physics data managers. In fact, EMODnet Physics portal distributes our dataset through a federated ERDDAP server's approach permitted by machine to machine interoperability. The practice of defining an url_metadata is advisable to manage more efficiently the numerous strings that are used for metadata information.

We decided to provide a Jupyter Notebook in Python (*Fратиanni and Frizzera, 2024*) that allows combining data and metadata into a single file to facilitate data reusability and be compliant with the FAIR data principles, as stated in the manuscript.

The notebook is available on a GitHub repository and published on Zenodo at <https://doi.org/10.5281/zenodo.13862792>. We have modified the annex C.6 to include this change.

Fратиanni, C., & Frizzera, P. (2024). REPROCESSED XBT 1999-2019: how to access data and metadata through ERDDAP (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.13862792>

- Attempts to download the full dataset as netcdf, csv or txt format failed due to the size of the dataset – is this a server-side issue? The error is:

```
Error {
  code=413;
  message="Payload Too Large: Your query produced too much data. Try to request less data.
[memory] The request needs more memory (80337 MB) than is ever safely available in this Java
setup (24072 MB). (TableWriterAll.cumulativeTable)";
}
```

Answer: We tried several times to download the entire dataset, both by selecting via browser interface the entire period and download as netcdf, csv and text format and via provided python script: we didn't have any issue. We are supposing that the encountered issue could be ascribed to incidental server load and traffic.

- In an attempt to get access to the full dataset, I tried using the url generated in ERDDAP in Matlab and Python.

```
http://oceano.bo.ingv.it/erddap/tabledap/REP_XBT_1999_2019_v2.nc?CALIB%2Cde  
pth%2CDEPTH_COR%2CDEPTH_COR_FLAGS_QC%2CDEPTH_COR_INT%2CDEPTH_COR_INT_SEADAT  
ANET_QC%2CDEPTH_FLAGS_QC%2CDEPTH_INT%2CDEPTH_INT_SEADATANET_QC%2CDEPTH_TEST  
_QC%2Clatitude%2Clongitude%2CPOSITION_SEADATANET_QC%2CSDN_BOT_DEPTH%2CSDN_C  
RUISE%2CSDN_EDMO_CODE%2CTEMPET01%2CTEMPET01_COR%2CTEMPET01_COR_FLAGS_QC%2CT  
EMPET01_COR_INT%2CTEMPET01_COR_INT_SEADATANET_QC%2CTEMPET01_FLAGS_QC%2CTEMP  
ET01_INT%2CTEMPET01_INT_SEADATANET_QC%2CTEMPET01_TEST_QC%2Ctime%2CTIME_SEAD  
ATANET_QC%2Ccruise_id%2Cprofile_id%2Curl_metadata&distinct()
```

If used as is, it crashed both Matlab and Python. After getting to the end of the paper, there is an example of python code to access the metadata for a single profile. A similar example of how to access the entire dataset with the url would be useful as the ERDDAP url above might need to be adjusted to work.

Answer: We have prepared a Jupyter Notebook (*Fратиanni, C. and Frizzera, P., 2024*), as suggested by the reviewer, to facilitate the access to the entire dataset and the recombination of data and metadata. The notebook is available on a GitHub repository and published on Zenodo at <https://doi.org/10.5281/zenodo.13862792>. We have modified Section 7 and annex C.6 to include this change.

Fратиanni, C., & Frizzera, P. (2024). REPROCESSED XBT 1999-2019: how to access data and metadata through ERDDAP (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.13862792

- Spelling errors in TEMPET01_TEST_QC variable attribute flag_meanings:

```
negative_vertical_gradient_at__iteration and positive_vertical_gradient_at__  
iteration
```

Also, the following flag_meanings do not match with the ones described in the paper and the flag_values are not included in the list (there are more flag_meanings than flag_values, as mentioned by Reviewer 2 in the first review).

```
negative_vertical_gradient_at_third_iteration  
positive_vertical_gradient_at_third_iteration
```

Answer: We corrected spelling mistakes and we also provided the corresponding flag meanings and flag values.

- In the first review I asked: Is the TEMPET01_TEST_QC variable additive? Your reply didn't address the issue (and I think Reviewer 2 asked the same question and described it better). Let me try again. There are several tests applied independently to each profile. If a single temperature fails more than one test, which exit value do you use? Or, do you record all the exit values for that data point and if you do, are they added together bit-wise as described in the netcdf conventions here: <https://cfconventions.org/cf-conventions/cf-conventions.html#flag-variable-flag-masks-flag-values-ex>. The way you are presenting the exit values does not meet the CF conventions and if more than one test is failed for a data point (as shown in your Figure 2 where both gross check and wire-stretch are failed for the same data points), how is that represented?

Answer: We thank the reviewer for stressing this point, we realize that the manuscript is not clearly describing how the QC tests are documented in the metadata. All the QC tests performed and their results (Table 2) are documented in three different ancillary variables: POSITION_SEADATANET_QC, DEPTH_TEST_QC, TEMPET01_TEST_QC. We slightly modified the text at the beginning of Section 4.1 and we added a column in Table 2.

None of the ancillary variables (POSITION_SEADATANET_QC, DEPTH_TEST_QC, TEMPET01_TEST_QC) is additive, we record each test exit value. We adopted the SeaDataNet convention and not the CF convention.

- The POSITION_SEADATANET_QC is defined by a single test (test 1 in Table 2).
- The DEPTH_TEST_QC contains the outcome of two tests, one based on GEBCO local bathymetry (test 2 in Table 2) and one based on the last good depth recorded by the operator (test 3 in Table 2). Since the GEBCO local bathymetry was often in disagreement with the operator information we decided to keep the output of test 3 in DEPTH_FLAGS_QC, the ancillary variable containing the resulting quality flags associated to each record in the depth profile (text has been added to Sec. 4.2).
- The TEMPET01_TEST_QC contains the exit values of six tests (tests 4-9) and Section 4.2 explains how we mapped the test exit values to the TEMPET01_FLAGS_QC. An example of the TEMPET01_TEST_QC (6 columns per a number of rows corresponding to the profile records) and the resulting TEMPET01_FLAGS_QC is reported below:

```
TEMP01_TEST_QC =
  49  50  48  48  48  49
  49  50  48  48  48  49
  49  49  48  48  48  49
  49  49  48  48  48  49
  49  49  48  48  48  49
  49  48  49  49  49  49
  49  48  49  49  49  49
  49  48  49  49  49  49
  49  48  49  49  49  49
  49  48  56  49  49  49
  49  48  49  49  49  49
  ...
```

```
TEMP01_FLAGS_QC =
  2
  2
  1
  1
  1
  1
  1
  1
  1
  1
  1
  1
  1
  1
  1
  ...
```

We slightly modified Section 4.2 to clarify the issues raised by the reviewer.