

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.

Manuscript specific comments:

1. 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.
2. Line 375 – 379: Position on land check. I asked about this in the previous review and your response was: *There are not profiles are 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.
3. Figure 10: 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.

Data comments:

1. 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.

```
// global attributes:  
    :area = "Mediterranean Sea" ;  
    :bathymetric_information = "102.0m" ;  
    :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/0000-0003-1283-2798, 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" ;
:Easternmost_Easting = 9.1445 ; Repeat of the geospatial

attributes

:fall_rate_equation_Coeff_1 = "6.301m s⁻¹" ;
:fall_rate_equation_Coeff_2 = "-0.00216m s⁻²" ;
:family_code = "XB" ;
:family_label = "vessel" ;
:featureType = "Profile" ;
:geospatial_lat_max = 44.2966 ;
:geospatial_lat_min = 43.8832 ;
:geospatial_lat_units = "degrees_north" ;
:geospatial_lon_max = 9.1445 ;
:geospatial_lon_min = 8.8679 ;
:geospatial_lon_units = "degrees east" ;
:geospatial_vertical_max = 550.59f ;
:geospatial_vertical_min = 0.f ;
:geospatial_vertical_positive = "down" ;
:geospatial_vertical_units = "m" ;
:history = "2024-07-08T02:56:09Z (local files)\n",
"2024-07-08T02:56:09Z
http://oceano.bo.ingv.it/erddap/taledap/REP_XBT_1999_2019_v2.nc?CALIB%2Cdepth%2CDEPTH_COR%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_C_RUISE%2CSDN_EDMO_CODE%2CTEMPET01%2CTEMPET01_COR%2CTEMPET01_COR_FLAGS_QC%2CTEMPET01_COR_INT%2CTEMPET01_COR_INT_SEADATANET_QC%2CTEMPET01_FLAGS_QC%2CTEMPET01_INT%2CTEMPET01_INT_SEADATANET_QC%2CTEMPET01_TEST_QC%2ctime%2CTIME_SEADATANET_QC%2Ccruise_id%2Cprofile_id%2Curl_metadata&time%3E=2019-09-21T00%3A00%3A00Z&time%3C=2019-09-28T00%3A21%3A13Z" ;
:ices_platform_code = "48AA" ;
:id = "T0_M1947" ;
:IMO_number = "8642751" ;
:infoUrl = "https://www.enea.it/it//,http://www.ingv.it" ;
:institution = "ENEA Centro Ricerche Ambiente Marino - La Spezia, Istituto Nazionale di Geofisica e Vulcanologia, Sede di Bologna (INGV)" ;
:institution_edmo_code = "136, 251" ;
:keywords = "1999-2019, ambiente, applied, applying, barker, bathymetric, bathymetry, bathythermograph, below, bologna, 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 > Bathymetry/Seafloor

Topography > Bathymetry, Earth Science > Oceans > Ocean Temperature > Water Temperature, earth science>oceans>ocean temperature>ocean temperature profiles, enea, european, expendible, flag, flags, floor, geofisica, grouping, ingv, interpolated, istituto, label, latitude, levels, ligurian, longitude, MACMAP, mapping, marine, marino, mcdougall, measurement, method, nazionale, ocean, oceans, organizations, original, partner, POSITION_SEADATANET_QC, profile, profile_id, quality, rep_xbt_1999_2019_v2, reprocessed, resolution, resulting, ricerche, science, SDN_BOT_DEPTH, SDN_CRUISE, SDN_EDMO_CODE, sea, sea_floor_depth_below_sea_surface, sea_water_temperature, seadatanet, seafloor, seas, seawater, sede, 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, thyrrhenian, time, TIME_SEADATANET_QC, topography, using, vertical, vulcanologia, water, xbt"

```
;  
:keywords_vocabulary = "GCMD Science Keywords" ;  
:last good depth according to operator = "92.78m" ;  
:last_latitude_observation = "44.2966" ;  
:last_longitude_observation = "9.1445" ;  
:launching_height = "05m" ;  
:license = "Attribution (CC BY 4.0)" ;  
:max_acquisition_depth = "300m" ;  
:max_recorded_depth = "292m" ;  
:netcdf_version = "netCDF-4 classic model" ;  
:Northernmost_Northing = 44.2966 ; Repeat of the geospatial
```

attributes

```
:pi_name = "Franco Reseghetti" ;  
: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" ;  
:sourceUrl = "(local files)" ;  
:Southernmost_Northing = 43.8832 ; Repeat of the geospatial
```

attributes

```
: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 (Expendible  
Bathythermograph (XBT)) probes sampled and managed by ENEA S.Teresa Centre  
since September 1999 in the Ligurian and Thyrrhenian 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; XCTD = 0.02 deg  
C (Table 2 from Cowley R et al., 2021  
https://doi.org/10.3389/fmars.2021.689695)" ;  
:time_coverage_end = "2019-09-28T00:21:13Z" ;  
:time_coverage_start = "2019-09-27T20:00:22Z" ;
```

```

        :title = "Reprocessed XBT dataset (V2) in the Ligurian and
Thyrrhenian Seas (1999-2019)" ;
        :update_interval = "void" ; Is this redundant, if so
remove?
        :Westernmost_Easting = 8.8679 ; Repeat of the geospatial
attributes
        :wmo_inst_type = "061" ;
        :wmo_platform_code = "IGMA" ;

```

2. 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.
3. 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.
4. 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)";
}

```

5. 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.

6. Spelling errors in TEMPET01_TEST_QC variable attribute flag_meanings:

negative_vertical_gradient_at_firts_iteration and
positive_vertical_gradient_at_firts_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

7. 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?