

EN.4.2.2 Product User Guide

V1.0

Rachel Killick (rachel.killick@metoffice.gov.uk)

Version	Description	Date of Creation
1.0	First EN4 Product User Guide	25 th August 2021

This version number will be incremented when changes are made to the documentation for additions or EN4 version updates.

CONTENTS

1	Introduction	4
2	Using EN4	4
2.1	How do I access the Data?	4
2.2	How do I use the data?	5
2.3	How do I cite the data?	5
2.4	What can I do with the Data?	6
2.5	What variables are in the data?	6
2.6	How often are the data updated?	6
2.7	What should I do if I have a problem with EN4?	6
2.8	Basic data overview	6
3	Public Data Files	10
3.1	Profile files	10
3.2	Analysis Files	14
4	XBT and MBT corrections.....	16
4.1	Default corrections	16
4.2	Levitus et al., 2009 corrections	16
4.3	Gouretski corrections.....	17
4.4	Cowley et al., 2013 corrections.....	18
4.5	Cheng et al., 2014 corrections	19
5	Example Code.....	19
5.1	Reading in EN4 profiles and plotting their locations	19
5.2	Merging monthly EN4 analysis files into an annual file	21
5.3	Creating a time series of a single analysis location.....	21
5.4	Plotting an analysis field and the observation weights	24
5.5	Plotting a salinity profile	27
5.6	Selecting only profiles that pass QC.....	28
5.7	Plotting profiles within a grid box and flagging levels that have failed qc	30
5.8	Identifying which QC checks have been failed.....	31
5.9	Working with the INST_REFERENCE and PROJECT_NAME variables.....	32
5.9.1	Finding delayed mode Argo profiles	32
5.9.2	Finding all XBT profiles and their types.....	33
6	FAQs	35
6.1	How does EN4 compare with other datasets?	35
6.2	What do the file names mean?	35

6.3	How does EN.4.2.2 differ from its predecessor?	36
6.4	How can I access older versions of the EN dataset?	37
6.5	Can I use EN4 for trend analysis?	37
6.6	Is X specific data source in EN4?	37
6.7	How do you deal with possible duplicates?	37
6.8	Do you include quality control flags from your input sources?	37
6.9	How do you deal with Argo data with a salinity drift?	37
6.10	Which instruments are used in EN4?	38
6.11	Which instruments are bias corrected in EN4?	38
6.12	Which set of bias corrections should I use for my application?	38
6.13	What is the reference pressure for Potential Temperature?	38
6.14	What are the units of the variables in EN4?	38
6.15	How is potential temperature calculated when no salinity is present?	38
6.16	How do I find the depth levels of the EN4 analyses?	38
6.17	How many levels are in the EN4 profile files?	38
6.18	What data conventions does EN4 follow?	39
6.19	What are the terms and conditions of using EN4?	39
6.20	Why do the analysis files start at -83N?	39
6.21	Are there uncertainties in EN4?	39
6.22	What is the precision of the data in EN4?	39
6.23	Where can I find more information about EN4?	39
6.24	How can I give feedback?	40
7	Glossary of acronyms and shorthand	40
8	References and Acknowledgements	40
	Appendix A	42

1 INTRODUCTION

EN.4.2.2, henceforth EN4, is the latest version in the 'EN' series of data sets from the Met Office Hadley Centre. It provides subsurface temperature and salinity data in both quality-controlled profile and objective analysis form from 1900 to the present day. Files are made available in netCDF format, with all profiles from the same month grouped together and an objective analysis file for each month. The data base is updated on a best endeavours basis, with approximately half a month's delay.

The data in EN4 come from four sources: Argo, the Arctic Synoptic Basin-wide Oceanography (ASBO) project, the Global Temperature and Salinity Profile Programme (GTSP), and the World Ocean Database 2018 (WOD18). Argo and GTSP data are collected from their processing centres every month, whilst the WOD18 data were taken as a single download in October 2020 and will not be updated until a new version of EN4 is created.

Data in EN4 come with bias adjustments for Mechanical BathyThermograph (MBT) and eXpendable BathyThermograph (XBT) data. There are four ensemble members in EN4, made from four different combinations of MBT and XBT bias adjustment schemes:

- Levitus et al., 2009 MBT and XBT corrections (denoted l09 in the file names),
- Gouretski and Cheng, 2020 MBT, and Gouretski and Reseghetti, 2010 XBT corrections (denoted g10),
- Levitus et al., 2009 MBT, and Cowley et al., 2013 XBT corrections (denoted c13),
- Gouretski and Cheng, 2020, MBT and Cheng et al., 2014 XBT corrections (denoted c14).

We encourage users to explore the impact these schemes have on their outputs. For more information on these bias correction schemes see [section 4](#).

2 USING EN4

2.1 HOW DO I ACCESS THE DATA?

EN4 files are available from <https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-2.html>. Announcements relating to any changes to the dataset will be posted on the download page, so we encourage users to check it before each download to ensure the most up to date information is being used.

Files are grouped by year, by ensemble member and by data type into .zip files. For example, to get all twelve monthly profile files in 2020 with Levitus et al., 2009 corrections applied you would need to download the EN.4.2.2.profiles.l09.2020.zip file.

There is guidance on the download page on the use of wget to download the entirety of one ensemble member. This relies on a .txt file with all the files of interest listed.

If you use the EN4 data in research, even without publication, we would be grateful if you could [let us know](#) for our records. If you would like to be added to our EN4 mailing list, which provides notices of version updates and bug fixes please also let us know.

2.2 HOW DO I USE THE DATA?

The EN4 data are made available in netCDF format. This is a platform-independent, self-describing binary format which can be worked with using a number of tools and programming languages. UCAR maintains a list of tools for working with such files¹. A selection of tools mentioned in this user guide or used in EN4 creation is:

- [ncdump](#): provided with the NetCDF library, produces a text rendering of a NetCDF file.
- [Climate Data Operators](#) (CDO)s: a set of command line utilities for performing operations on NetCDF files including concatenation, editing and mathematics.
- The Python [netCDF4](#) package.
- The Python [iris](#) package.

2.3 HOW DO I CITE THE DATA?

When working with EN4 data please cite the EN4 paper, Good et al., 2013, and the full version number you are using, as well as the bias correction schemes you have used. E.g. "EN.4.2.2 (Good et al., 2013) ensemble members using Levitus et al., 2009 and Cowley et al., 2013 bathythermograph corrections were used in this work".

Please also include an acknowledgement/ reference of the input sources to the data that you have used. These are as follows for the different datasets:

Argo: "The Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (<https://argo.ucsd.edu>, <https://www.ocean-ops.org>). The Argo Program is part of the Global Ocean Observing System"

ASBO sources:

- Hydrobase - Curry (2001).
- Beaufort gyre experiment - The data were collected and made available by the Beaufort Gyre Exploration Program based at the Woods Hole Oceanographic Institution (<https://www2.whoi.edu/site/beaufortgyre/>).
- North Pole Environmental Observatory (NPEO; <http://psc.apl.washington.edu/northpole>) - NSF Grants OPP-9910305, OPP-0352754, and ARC-0634226.
- Freshwater switchyard of the Arctic project - M. Steele, W. Smethie, and P. Schlosser, CTD data from the Arctic Switchyard Project, Office of Polar Programs, National Science Foundation: <http://psc.apl.washington.edu/switchyard/>.
- Nansen and Amundsen Basins Observational System (NABOS), Canadian Basin Observational System (CABOS) (<http://nabos.iarc.uaf.edu/>).
- Physical and Chemical Properties from Selected Expeditions in the Arctic Ocean (Swift 2002; [EOL data archive -- Physical and Chemical Properties from Selected Expeditions in the Arctic Ocean \(ucar.edu\)](#)).

GTSP: Sun, C. and Co-Authors (2010). "The Data Management System for the Global Temperature and Salinity Profile Programme" in Proceedings of OceanObs.09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21-25 September 2009, Hall, J., Harrison, D.E., & Stammer, D., Eds., ESA Publication WPP-306, [doi:10.5270/OceanObs09.cwp.86](https://doi.org/10.5270/OceanObs09.cwp.86)

¹ <https://www.unidata.ucar.edu/software/netcdf/software.html>

WOD18: Boyer, T.P., O.K. Baranova, C. Coleman, H.E. Garcia, A. Grodsky, R.A. Locarnini, A.V. Mishonov, C.R. Paver, J.R. Reagan, D. Seidov, I.V. Smolyar, K. Weathers, M.M. Zweng (2018): World Ocean Database 2018. A.V. Mishonov, Technical Ed., NOAA Atlas NESDIS 87.
https://www.ncei.noaa.gov/sites/default/files/2020-04/wod_intro_0.pdf

2.4 WHAT CAN I DO WITH THE DATA?

We encourage you to make use of all of the data relevant to your specific application. Please note the licence restrictions on the website, replicated in all of the netCDF files. EN4 is served under a [non-commercial government licence](#).

See [section 5](#) for some simple examples of working with EN4 data.

If you use the data, please [get in touch](#) and let us know how you have used them and provide feedback on their usability, this will help us to continue to develop this product.

2.5 WHAT VARIABLES ARE IN THE DATA?

EN4 is a subsurface temperature and salinity data set, focusing on potential temperature² and practical salinity³, but with in situ temperatures included in the profile files as well. It does not include any biogeochemical observations at the current time. For information on the exact variables included and their QC flags please see [section 3](#).

2.6 HOW OFTEN ARE THE DATA UPDATED?

EN4 files are updated on a best-endeavours basis usually between the 10th and 20th of the month. A monthly release will provide a final (.f.) file for two months prior and a preliminary (.p.) file for the previous month. Once a final file has been created it will not usually be changed until a new version of the data set is released, which will then be indicated by a change in version number. If an error is found meriting a change to the .f. files without a change in version number this will be clearly indicated on the EN4 website and a copy of the old files will be archived for traceability.

2.7 WHAT SHOULD I DO IF I HAVE A PROBLEM WITH EN4?

If you have a query about using the EN4 data, you can search our [FAQs](#) to see if your query is answered there. If it is not, or you require further assistance, then please [contact us](#).

2.8 BASIC DATA OVERVIEW

EN4 runs from 1900 to the present day, with approximately a one month lag. Figure 1 shows how the quantity of profiles increases over time and compares the profile counts in EN.4.2.2 and EN.4.2.1. Both EN4 versions show an increase in monthly profile counts in the 2000s, this is linked to the increasing deployment of Argo floats. From approximately 2004 onwards, EN.4.2.2 consistently includes a greater amount of data, this is owing to the more recent download of the source data to pick up additional profiles. The difference between EN.4.2.2 and EN.4.2.1 becomes most noticeable

² "Potential temperature is the temperature a parcel of air or sea water would have if moved adiabatically to sea level pressure" from <https://cfconventions.org/Data/cf-standard-names/28/build/cf-standard-name-table.html>.

³ "Practical Salinity, S_P, is defined on the Practical Salinity Scale of 1978 (PSS-78) and is calculated from the electrical conductivity of sea water (as well as temperature and pressure)". From <https://cfconventions.org/Data/cf-standard-names/28/build/cf-standard-name-table.html>.

after 2016, which was when the last major data download (from all sources) was done for EN.4.2.1. Likewise, there is a drop in the number of profiles in EN.4.2.2 in late 2020, as this is when WOD profiles stopped being ingested.

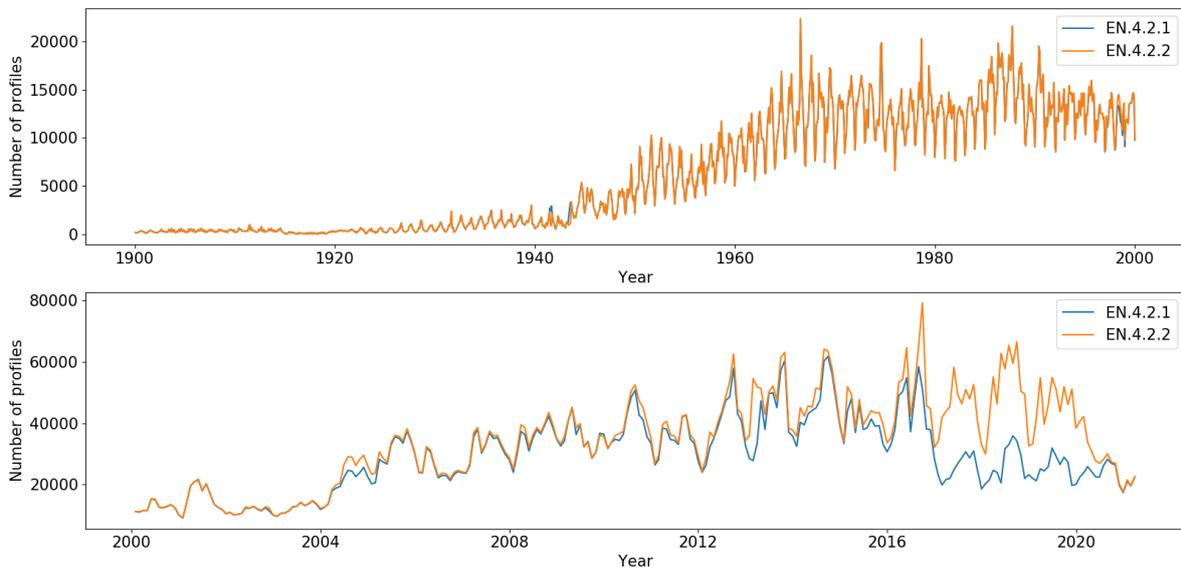


Figure 1: Total profile count in EN.4.2.2 compared to EN.4.2.1. Top: From Jan 1900 to Dec 1999. Bottom: From Jan 2000 to April 2021.

Figure 2 shows the number of profiles considered good (no whole profile QC checks failed) for temperature and salinity in EN.4.2.2. It shows that until the 1940s the two variables are similarly represented. With the advent of Mechanical BathyThermographs (MBTs) and eXpendable BathyThermographs (XBTs) in the late 1930s and late 1960s respectively, the prevalence of the two variables diverges as MBTs and XBTs do not measure salinity. Coverage becomes more similar with the introduction of Argo, though still with an offset.

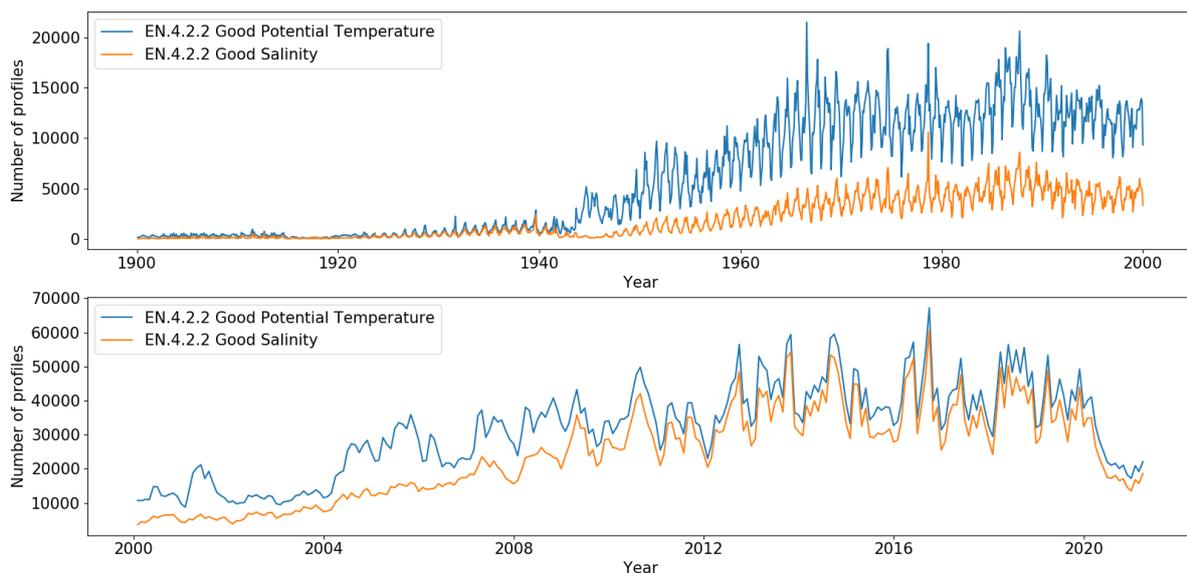


Figure 2: Total number of EN.4.2.2 profiles considered good for potential temperature (blue) and salinity (orange). Top: From Jan 1900 to Dec 1999. Bottom: From Jan 2000 to Apr 2021.

Figure 3 shows the increase in profile spatial and depth coverage over time. It also highlights again the discrepancy in coverage between temperature and salinity as well as the areas where coverage

could still be improved. For an approximate breakdown of the use of different instruments see the [HadIOD.1.2.0.0 user guide](#), although note that HadIOD.1.2.0.0 was built from earlier versions of EN4 and incorporates much surface data. It will, therefore, not have exactly the same coverage as EN.4.2.2.

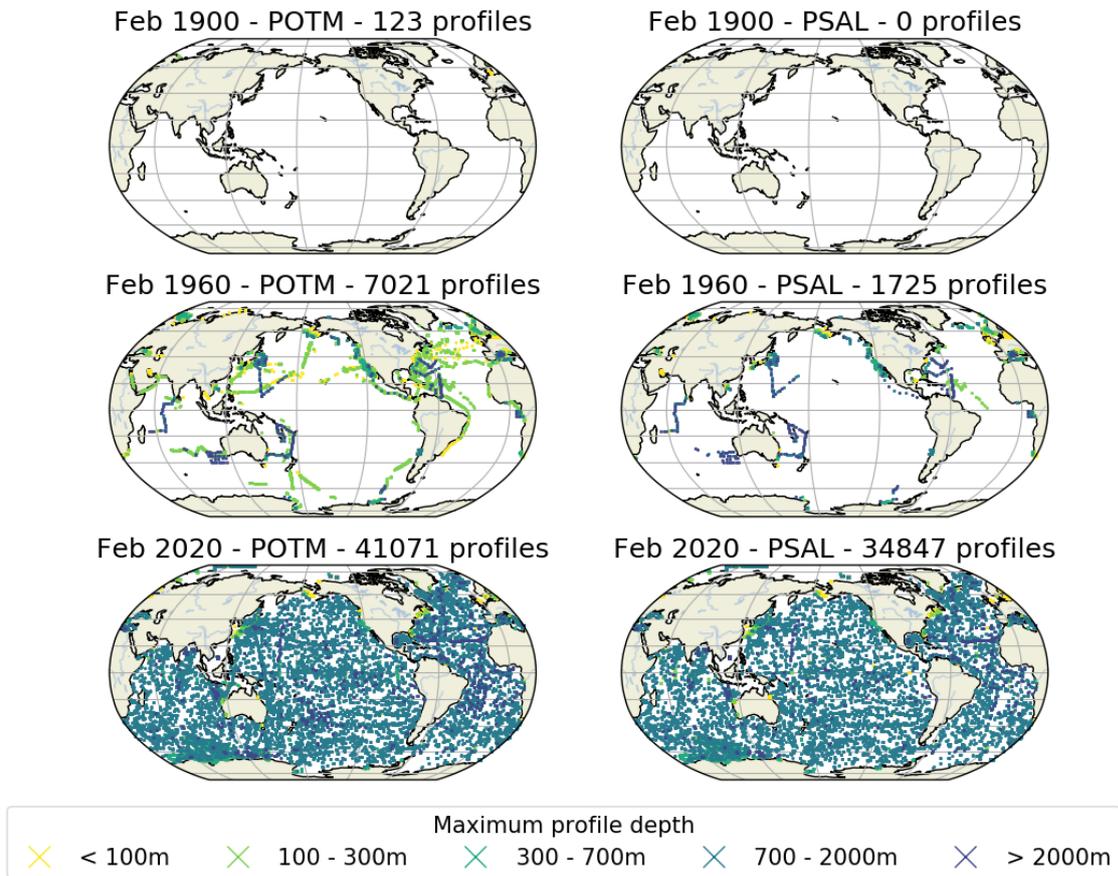


Figure 3: Change in profile coverage and depth over time for EN.4.2.2. POTM = potential temperature. PSAL = practical salinity.

Figure 4 shows the relative weighting of the observations to the background in the temperature and salinity analyses at three different levels. Where values are close to one it means the observations have closely influenced the analysis at that point, whilst a value of zero means there were no observations influencing the analysis at that point. For February 1900 the surface layer (down to 10m) is shown, for February 1960 the layer incorporating 300m is shown, and for February 2020 the layer down to almost 2000m is shown. This figure complements Figure 3 to illustrate the changing availability of observations over time.

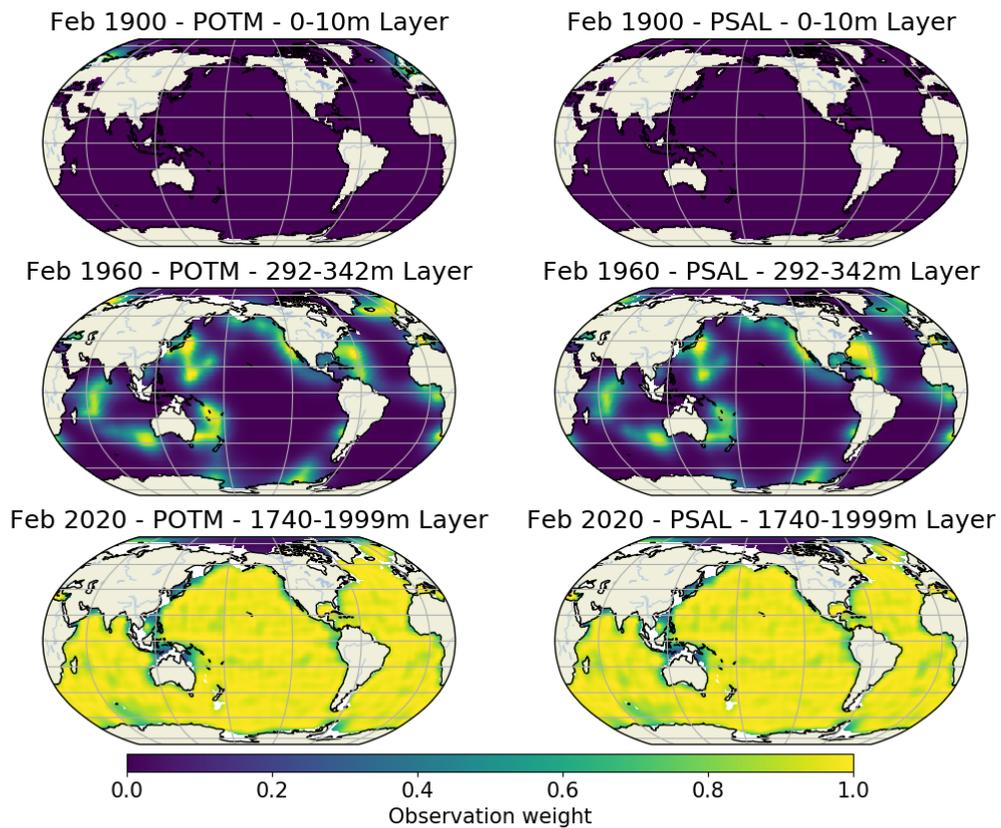


Figure 4: Observation weight fields for three months and three levels of the EN.4.2.2 analyses.

Figure 5 shows the impact of different bias corrections, and why they are necessary, using the example of 0 – 700m ocean heat content anomalies from 1950 to 2020 with a 1950-2020 climatology.

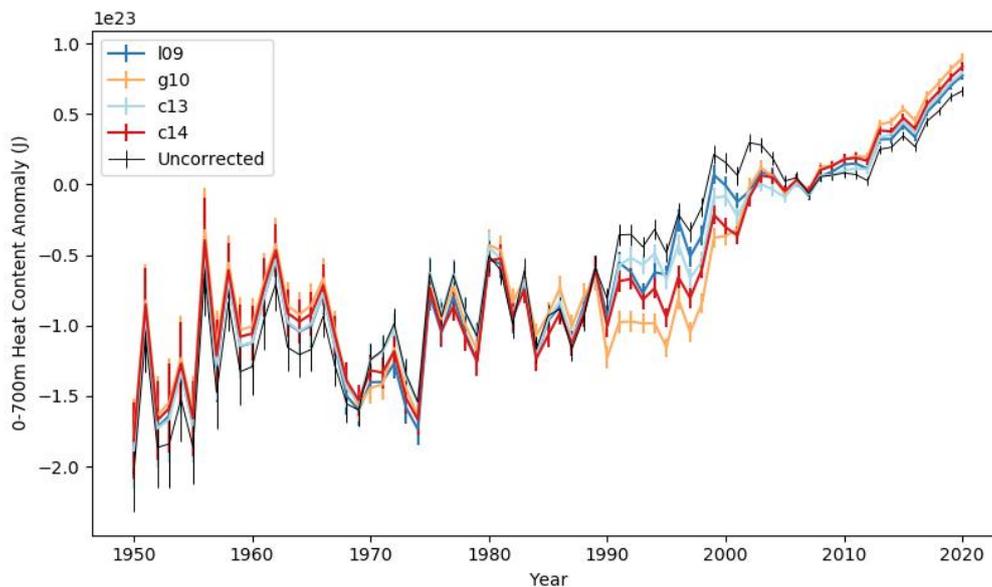


Figure 5: 0 – 700m global ocean heat content from 1950 to 2020. Blue = l09 corrections; orange = g10 corrections; pale blue = c13 corrections; red = c14 corrections; and black = no XBT/MBT bias correction scheme applied.

3 PUBLIC DATA FILES

EN4 is made up of monthly profile files and analysis files, available for public use, under the restrictions laid out in [the licence](#). There are four ensemble members for each of these file types; those with Levitus et al., 2009 XBT and MBT corrections (l09); Gouretski and Reseghetti, 2010 XBT and Gouretski and Cheng, 2020 MBT corrections (g10); Cowley et al., 2013 XBT and Levitus et al., 2009 MBT corrections (c13); and Cheng et al., 2014 and Gouretski and Cheng, 2020 MBT corrections (c14). We encourage users to explore the impact of different ensemble members on their application to ensure that uncertainty arising from necessary, but imperfect, bathythermograph bias corrections is represented.

3.1 PROFILE FILES

EN4 profile files contain all profiles from a given month grouped together in a single netCDF file. Within this file there are a number of variables as laid out in Tables 1 and A1. A single profile is a set of vertical observations at the same geographical location at the same time from the same instrument. All profile observations have undergone automatic quality control and we encourage users to make use of the QC flags provided.

Table 1: Key variables contained in the EN4 netCDF profile files. Other variables contained in these files are described in [Table A1](#).

NetCDF Variable	Description
Global Attributes	<i>Attributes describing the whole file.</i>
Licence	The licence under which these data can be used. This is a non-commercial government licence .
References	This gives a link to both the website and the paper, please see section 2.3 on how to cite the data in publications.
Dimensions	<i>These are the same dimension variables as in an Argo file, although the values given are specifically for EN4.</i>
N_PROF	The number of profiles in the file – not fixed.
N_LEVELS	Number of levels allowed in a profile – maximum of 400.
General Information	<i>These variables contain information about the whole profile file. They share the same variable names as those in Argo.</i>
REFERENCE_DATE_TIME	Date of reference for Julian days. This is always 19500101000000 in DATE_TIME format i.e. 1 st January 1950 and is necessary for calculating the JULD in human readable form.
General Information for each profile	<i>These are the variables that have an N_PROF dimension. Many share the same variable names as those in Argo, but the contents are different in places.</i>
INST_REFERENCE	The contents are dependent on the data source, but enable the user to learn more about the instrument used to make this profile.

	Argo	WMO code table 1770 ⁴ . The relevant subsection of which is in Reference Table 8 of the Argo Users' Manual.
	ASBO	Data are from bottles or CTDs and have a fixed INST_REFERENCE of 741.
	GTSP	WMO code table 1770. With additional entries of 101 if the inst_type is unknown, but a fall rate of A=6.472 is assumed and 102 if the inst_type is unknown, and no fall rate correction is required.
	WOD	10000*ProbeType + 10*InstType + CorrType. ProbeType is WOD secondary header 29 and gives a broad instrument type classification ⁵ ; InstType is a variable specific secondary header and gives the specific instrument used to take the measurement ⁶ ; CorrType is derived from WOD secondary header 54 ⁷ , taking entries for 0, 1 and 2 as given and replacing 103 by 3, 104 by 4 and -1 by 9 to ensure this variable is only one character long.
JULD		Julian day of the profile relative to REFERENCE_DATE_TIME (1 st January 1950).
LATITUDE		The latitude of the profile. In degrees North.
LONGITUDE		The longitude of the profile. In degrees East.
PLATFORM_NUMBER		This variable is helpful for tracing platforms.
	Argo	The WMO float identifier.
	ASBO	This is comprised of a two character country code, followed by the first character of the ship code followed by a five character cruise number.
	GTSP	This is the GTSP call sign (or csign).
	WOD	The first characters are the WOD cruise identifier (without the country letters at the start), followed by the final two characters of the WOD platform code (which is included in full in the PROJECT_NAME variable).
POSITION_QC		Quality on position. 1 for good and 4 for bad.
PROFILE_POTM_QC		Global quality flag of the potential temperature profile. 1 for good and 4 for bad.
PROFILE_PSal_QC		Global quality flag of the practical salinity profile. 1 for good and 4 for bad.
PROJECT_NAME		The first characters of this variable show the data source, one of Argo, ASBO, GTSP or WOD09 ⁸ . The characters following this are dependent on the data source. This variable is important for correct data acknowledgement and can be used for further data traceability.

⁴ <https://www.nodc.noaa.gov/gtspp/document/codetbls/wmocodes/table1770.html>.

⁵ [s_29_probe_type.pdf \(noaa.gov\)](#)

⁶ [v_5_instrument.pdf \(noaa.gov\)](#)

⁷ [s_54_needs_depth_fix.pdf \(noaa.gov\)](#)

⁸ World Ocean Database 2018 was used for this work, but an error in processing means that the project reference remained WOD09.

	Argo	Character 5 indicates whether these data are real-time (R), real-time adjusted (A) or delayed-mode (D).
	ASBO	Characters 6 and 7 indicate the instrument type from the GTSP type table ⁹ . This is followed by a blank character and then country, ship, cruise, and station numbers each separated by a blank character.
	GTSP	Characters 6 and 7 are as for ASBO. There is then a blank character followed by a two character code indicating the recorder type from WMO code table 4770 ¹⁰ .
	WOD	Characters 6 to 8 are the instrument type, and 9-10 are the country code ¹¹ and the next eight characters contain the platform code ¹² .
QC_FLAGS_PROFILES		This contains detailed information on why a profile has a good or bad PROFILE_<PARAMETER>_QC flag. Each bit within the value has a different meaning if set, the meanings of these bits are given in Table 2.
Measurements for each profile		<i>These are the variables that have an N_PROF and an N_LEVELS dimension.</i>
DEPH_CORRECTED		The depth of each profile level. These depths include any adjustments made during bias correction.
POTM_CORRECTED		The potential temperature of each profile level. These temperatures include any adjustments made during bias correction.
POTM_CORRECTED_QC		Quality on each POTM_CORRECTED level. 1 = good, 4 = bad, 0 = masked data and therefore no QC performed.
PSAL_CORRECTED		The practical salinity of each profile level.
PSAL_CORRECTED_QC		Quality on each PSAL_CORRECTED level. 1 = good, 4 = bad, 0 = masked data and therefore no QC performed.
QC_FLAGS_LEVELS		This contains detailed information on why a level has a good or bad <PARAMETER>_CORRECTED_QC flag. Each bit within the value has a different meaning if set, the meanings of these bits are given in Table 3.
TEMP		The in situ temperature of each profile level. These temperatures include any adjustments made during bias correction.
Bias correction scheme specific variables		<i>These are variables which are only present in specific ensemble members. Their values are missing if they are not applied to a profile e.g. it is not a bathythermograph or is of an 'uncorrectable' type. If they are present, they have already been applied and are given only for reference.</i>

⁹ <https://www.nodc.noaa.gov/gtspp/document/codetbls/gtsppcode.html#TYPE>.

¹⁰ https://library.wmo.int/doc_num.php?explnum_id=10235 (You will need to search for table 4770 within this document).

¹¹ <https://www.ncei.noaa.gov/access/world-ocean-database/CODES/country-list.html>

¹² https://www.ncei.noaa.gov/access/world-ocean-database/CODES/s_3_platform.html

CORRECTIONS_ALPHA	For the Cowley et al., c13, ensemble member. $1 - \text{CORRECTIONS_ALPHA}$ is the depth multiplier.
CORRECTIONS_BETA	For the Cowley et al., c13, ensemble member. CORRECTIONS_BETA is the value subtracted off the depth (after it has been multiplied by $1 - \text{CORRECTIONS_ALPHA}$).
CORRECTIONS_THERMAL	For the Cowley et al., c13, ensemble member. This is the value subtracted off the temperatures.
LEVITUS_MBT_TEMP_CORRECTION	For the Cowley et al., c13 and Levitus et al., l09 ensemble members. This is the value added to the MBT temperatures.
LEVITUS_XBT_TEMP_CORRECTION	For the Levitus et al., l09, ensemble member. This is the value added to the XBT temperatures.
Depth_corrections	For the Gouretski et al., g10 and Cheng et al., c14, ensemble members. These are the values subtracted from the depths.
Thermal_corrections	For the Gouretski et al., g10 and Cheng et al., c14, ensemble members. These are the values subtracted off the temperatures.

Table 2: Meaning of bits in QC_FLAGS_PROFILES. For more information on the QC tests see Good et al., 2013.

Bit	Meaning if set
0	Temperature profile was rejected.
1	Salinity profile was rejected.
2	Indicates there are other profiles within 0.2 degrees of latitude and longitude and 1 hour in time that appear to be of higher quality.
3	Temperature and salinity profile rejected by the track check.
4	Temperature and salinity profile rejected by the vertical stability check.
5	Temperature and salinity profile rejected because the profile is on the altimetry QC suspect list.
6	Temperature and salinity profile rejected because the profile appears to be on land.
8	Temperature profile rejected because it is at exactly zero degrees latitude and longitude.
9	Temperature profile rejected because it is on the Argo grey list.
10	Temperature profile rejected because it is on the EN4 reject list (which includes profiles which are considered bad after human inspection).
11	Temperature profile rejected by the vertical check, which looks for spikes etc.
12	Temperature profile rejected because there were no background values for this profile.
13	Temperature profile rejected because over half of its levels were rejected.
16	Salinity profile rejected because it is at exactly zero degrees latitude and longitude.
17	Salinity profile rejected because it is on the Argo grey list.
18	Salinity profile rejected because it is on the EN4 reject list.
19	Salinity profile rejected by the vertical check.
20	Salinity profile rejected because there were no background values for this profile.
21	Salinity profile rejected because over half of its levels were rejected.
24	Information flag to say that a correction has been applied to the XBT depths in the pre-processing (i.e., before any XBT correction schemes unique to the ensemble members have been applied).
25	Information flag to say that this profile is a 'superob', an average of multiple observations, for moored buoys only.

Table 3: Meaning of bits in QC_FLAGS_LEVELS.

Bit	Meaning if set
0	Temperature level was rejected.
1	Salinity level was rejected.
2	Temperature and salinity levels rejected because they failed the vertical stability check.
3	Temperature and salinity levels rejected because they failed the increasing depth check.
8	Temperature level was rejected because it is a bathythermograph measurement shallower than 4m or deeper than 950m; this flag is also set if latitude and longitude are at exactly zero degrees.
9	Temperature level rejected because it was rejected by Argo QC procedures.
10	Temperature level was rejected because it was outside a reasonable range and was set to missing data.
11	Temperature level was rejected because it was on the EN4 reject list.
12	Temperature level was rejected by the vertical check.
13	Temperature level was rejected because there was no background value available.
14	Temperature level was rejected by the Bayesian background check performed on reported levels.
15	Temperature level was rejected by the Bayesian and buddy checks performed on standard levels.
16	Temperature level was reinstated after rejection by the Bayesian and buddy checks performed on standard levels.
21	Salinity level was rejected because it was rejected by Argo QC procedures.
22	Salinity level was rejected because it was outside of a reasonable range and was set to missing data.
23	Salinity level was rejected because it was on the EN4 reject list.
24	Salinity level was rejected by the vertical check.
25	Salinity level was rejected because no background value was available.
26	Salinity level was rejected by the Bayesian background check performed on reported levels.
27	Salinity level was rejected by the Bayesian and buddy checks performed on standard levels.
28	Salinity level was reinstated after rejection by the Bayesian and buddy checks performed on standard levels.
29	Salinity level was rejected because it failed the waterfall check.

Example code to extract which bits have been set in QC_FLAGS_PROFILES and QC_FLAGS_LEVELS is given in [section 5](#).

3.2 ANALYSIS FILES

EN4 analysis files are objective analyses formed using the analysis correction scheme. They combine the quality controlled profiles with a background field, which is a persistence based forecast of the previous month's objective analysis, incorporating the long term, 1971-2000, climatology. Only profiles/ profile levels that have passed the QC checks will contribute to the analyses. In the long term absence of observations, the analyses will relax back to climatology, we therefore recommend using the observation weights fields in the analyses to determine the relative influence of observations and climatology before using analysis fields, especially for long term trend analysis.

EN4 analysis files are monthly mean files on a 1 by 1 degree horizontal grid over 42 irregularly spaced depth levels. The variables included in this file are given in Table 4.

Table 4: Contents of the EN4 analysis files. Additional global variables that are not necessary for use of the analysis files but may be of interest to users investigating their full contents are given in [Table A2](#).

NetCDF Variable	Description
Global Attributes	<i>Attributes describing the whole file.</i>
Licence	The licence under which these data can be used. This is a non-commercial government licence .
References	This gives a link to both the website and the paper, please see section 2.3 on how to cite the data in publications.
Dimensions	<i>These are the dimensions in the analysis files.</i>
Depth	There are always 42 depth levels in EN4 analysis files.
Lat	Always 173, to account for the most southerly latitudes, where there is no ocean, being excluded.
Lon	360.
Time	Each analysis file represents the conditions in one month.
Bnds	There are two bnds or bounds variables, depth_bnds and time_bnds. These give the boundaries of the file (see below).
Information variables	<i>These are the variables that will be defined for each grid point.</i>
Depth	The depth, in metres, of the approximate midpoint of the analysis level. It is the approximate midpoint because of historical re-gridding.
Depth_bnds	The boundaries of the depth layers. E.g., The first layer represents the depths 0m to 10.0475m.
Lat	The central latitude of the grid point (in degrees north).
Lon	The central longitude of the grid point (in degrees east).
Salinity	The salinity of the grid box.
Salinity_observation_weights	The relative weighting given to the observations compared to the background in the salinity analysis. If the number is close to one it means that observations have closely influenced the analysis while a zero indicates that no observations were available during the month in/near that location.
Salinity_uncertainty	The analysis error standard deviation of the salinity analysis.

Temperature	The potential temperature of the grid box. In Kelvin (although some software will automatically subtract off the 273.15 degree offset to convert this to Celsius).
Temperature_observation_weights	The relative weighting given to the observations compared to the background in the potential temperature analysis.
Temperature_uncertainty	The analysis error standard deviation of the potential temperature analysis.
Time	The middle of the month that was analysed. In days since 1 st Jan 1800.
Time_bnds	The boundaries (beginning and end times) of the month being analysed. In days since 1 st Jan 1800.

4 XBT AND MBT CORRECTIONS

All EN.4.2.2 data come with XBT and MBT bias corrections applied. The bias corrections applied determine the code in the ensemble member file name, one of l09, g10, c13 or c14, as stated in the [Introduction](#). This section gives an overview of each of these ensemble members.

4.1 DEFAULT CORRECTIONS

Before any additional XBT corrections are applied in EN4 the recommended Hanawa et al., 1995 adjustment is applied for all T4, T6, T7 and Deep Blue probes that have only had manufacturer fall rate equations applied (this is a depth multiplier of 1.0336). The recommended Kizu et al., 2005 corrections for TSK T5s are also applied at this stage (this is a depth dependent multiplication factor for reported depths, see table 2 of Kizu et al., 2005).

References: K. Hanawa et al., 1995. A new depth-time equation for Sippican or TSK T-7, T-6 and T-4 expendable bathythermographs (XBT). *Deep-Sea Res.* 42, 1423–1451.

S. Kizu et al., 2005: Inter-manufacturer difference and temperature dependency of the fall-rate of T-5 expendable bathythermograph. *Journal of Oceanography.* 61 (5), 905–912.

4.2 LEVITUS ET AL., 2009 CORRECTIONS

Corrections available: This scheme contains corrections for both MBTs and XBTs. These corrections are year dependent, and depth dependent, but they are not probe type dependent. There is one exception to this, XBTs that are T5s (or suspected T5s) are not corrected. Suspected T5s are probes with unknown type that have a maximum depth greater than 900m.

Coverage: Year specific Levitus corrections were available for 1947 to 1992 for MBTs and 1969 to 2015 for XBTs. Outside these year ranges, in EN4, the correction values from the closest year are implemented. For example, an MBT in 1940 would receive the MBT corrections given for 1947 and an XBT in 2021 would receive the XBT corrections for 2015.

File identifier in EN4: In EN4 files with Levitus XBT and MBT corrections contain the string 'l09' in their names.

Last updated: The Levitus corrections in EN4 were provided in July 2019 and remain the most up to date.

Reference: S. Levitus et al., 2009: Global Ocean Heat Content 1955-2008 in light of recently revealed instrumentation problems. *Geophysical Research Letters*, 36, L07608, doi: <http://dx.doi.org/10.1029/2008GL037155>

4.3 GOURETSKI CORRECTIONS

Corrections available: Up to and including EN.4.2.1 both XBTs and MBTs were corrected using an updated version of the Gouretski and Reseghetti, 2010 corrections. From EN.4.2.2 onwards, XBTs were still corrected using the updated Gouretski and Reseghetti, 2010 values, whilst MBTs were corrected using the Gouretski and Cheng, 2020 corrections.

XBT corrections: The Gouretski and Reseghetti, 2010 XBT corrections provide some granularity in the corrections applied. They distinguish between XBT manufacturers; Sippican, TSK and Sparton and also between some probe types. Only probes known or suspected to be manufactured by Sippican are corrected by this scheme. If a probe's manufacturer is unknown, but it does not come from a 'TSK country' then it is assumed to be a Sippican probe and is corrected based on its maximum depth. 'TSK countries' are those known to have historically used TSK probes, these countries are China, Japan, Taiwan, South Korea and Thailand. As Sparton were a smaller XBT manufacturer, no unknown probes are assumed to be Spartons by this correction scheme.

There are both depth and temperature corrections available in the Gouretski and Reseghetti XBT corrections. Separate corrections are available for Sippican T4/T6 probes, Sippican T5 probes, Sippican T7/Deep Blue probes and Sippican T10 probes. All other probe types (which are a small proportion of the data) e.g., T11s and Fast Deeps are not corrected by this XBT correction scheme. For the probe types to be corrected, the Hanawa corrections are undone first to ensure we are starting with data of the same form that were used to derive the Gouretski and Reseghetti corrections.

MBT corrections: The Gouretski and Cheng, 2020 MBT corrections are available to be applied to all MBTs. They consist of a set of depth, year and country dependent depth correction coefficients and a country and year dependent temperature correction. The country groupings are USA, Russia, Japan, Canada, Great Britain, and 'other countries'. The country that is recorded as launching the MBT is assumed to be the country that the MBT originated from.

The published depth corrections were depth dependent and had the potential to become very large at deeper depths. Therefore, below 200m, the depth correction applied in EN4 was that derived for the 200m depth. This recommendation was made by one of the scheme's developers (personal communication).

Coverage: Not all corrections available in the Gouretski and Cheng, 2020 paper were used. The only corrections used were those corrections where at least 2000 profiles were available at their derivation and the suggested depth corrections did not exceed 10% of the recorded depth or 10m, whichever was greater. Where these conditions were not met one of the following was done:

- For year-country pairings earlier (later) than the first (last) 'good' year: the corrections from the closest available 'good' year for that country were used.
- For year-country pairings where there were 'good' years either side: the corrections from the 'good' years either side were linearly interpolated.

For the depth corrections, correction coefficients, instead of correction values, were provided, and it was these coefficients that were interpolated/ repeated.

Corrections designed for one country were never used to infill gaps in another country's corrections.

File identifier in EN4: In EN4, files with Gouretski XBT and MBT corrections contain the string 'g10' in their names.

Last updated: The Gouretski and Reseghetti XBT corrections used were last updated by the authors in 2012. The MBT corrections are new and were last updated at the start of 2021.

References: Viktor Gouretski and Franco Reseghetti, 2010: On depth and temperature biases in bathythermograph data: development of a new correction scheme based on analysis of a global ocean database. *Deep-sea Research*, 57, 6, doi: <http://dx.doi.org/10.1016/j.dsr.2010.03.11>

Viktor Gouretski and Lijing Cheng, 2020: Correction for Systematic Errors in the Global Dataset of Temperature Profiles from Mechanical Bathythermographs. *Journal of Atmospheric and Oceanic Technology*, 37, 841 – 855, doi: 10.1175/JTECH-D-19-0205.1

4.4 COWLEY ET AL., 2013 CORRECTIONS

Cowley et al., provide only XBT corrections in their 2013 paper. In EN4 these XBT corrections are paired with the Levitus et al., 2009 MBT corrections to ensure that both bathythermograph types are still corrected.

These XBT corrections are probe and manufacturer dependent. Separate corrections are provided for Sippican T4/T6, Sippican T7/Deep Blue, TSK T4/ T6 and TSK T7/Deep Blue probes. T10s and T11s are always corrected with the Sippican T4/T6 corrections, while Fast Deeps are corrected with the T7/Deep Blue corrections appropriate to their manufacturer. T5s are not corrected by this scheme, nor are probes of unknown type going deeper than 1005m, which this scheme uses as the T5 depth cut-off. AXBTs (aerial XBTs) are also not corrected by this scheme. Sparton T4, T6, T7, T10 and Deep Blue probes are corrected using the appropriate Sippican corrections. If the manufacturer of a probe is not known, then it is assigned based on the same use of TSK countries and non-TSK countries as defined in the Gouretski XBT corrections. If the probe type is unknown, then it is assigned based on the maximum depth of the profile and its year. Corrections are provided for both temperature and depth.

Coverage: In years where T4/T6 corrections were not available the T7/DB corrections were used and vice versa. In years where Sippican correction values were not available the TSK values were used and vice versa. Outside of the years that any corrections were provided for, the corrections from the nearest available year were used. Note that the Cowley scheme does check the date of an XBT profile against the first date of manufacture for that type of XBT, if it is too early then the profile is not corrected.

File identifier in EN4: In EN4, files with Levitus MBT corrections and Cowley XBT corrections contain the string 'c13' in their names.

Last updated: The Cowley et al corrections were confirmed as the most up to date in April 2020.

Reference: R. Cowley, S. Wijffels, L. Cheng, T. Boyer, and S. Kizu, 2013: Biases in Expendable Bathythermograph Data: A new view based on historical side-by-side comparisons. *Journal of Atmospheric and Oceanic Technology*. 30. 1195 – 1225. Doi: 10.1175/JTECH-D-12-00127.1

4.5 CHENG ET AL., 2014 CORRECTIONS

Cheng et al., provide only XBT corrections in their paper and these are paired with the Gouretski and Cheng, 2020 MBT corrections in EN4.

There are nine groups of XBT corrections in Cheng et al., providing both temperature and depth corrections which are time, temperature, and probe type dependent. These nine groupings are G1 – Sippican T7/Deep Blue; G2 - Deep probe of unknown type and make; G3 – Sippican T4/T6; G4 - Shallow probe of unknown type and make; G5 – Sippican/TSK T10s; G6 – Sippican T5; G7 – TSK T4/T6; G8 – TSK T5 and G9 – TSK T7/Deep blue. Sparton XBTs are not corrected by this scheme.

Coverage: Corrections are available from 1966 to 2016, outside these years the correction coefficients from the closest available year (for each group) were used. For other correction schemes the given yearly coefficient/ correction values were interpolated to the exact time of the profile. This was not done in this scheme at the request of the scheme’s lead author, instead all profiles of the same type in a calendar year had their corrections derived using that year’s coefficients, with no weighting toward the coefficients of neighbouring years.

File identifier in EN4: In EN4 files with Gouretski and Cheng, 2020, MBT corrections and Cheng XBT corrections contain the string ‘c14’ in their names.

Last updated: The Cheng et al corrections were confirmed as the most up to date in 2020.

Reference: L. Cheng, J. Zhu, R. Cowley, T. Boyer, and S. Wijffels, 2014: Time, Probe Type and Temperature Variable Bias Corrections to Historical Expendable Bathythermograph Observations. *Journal of Atmospheric and Oceanic Technology*. 31. 1793 – 1825. Doi: 10.1175/JTECH-D-13-00197.1.

5 EXAMPLE CODE

All these examples make use of Python 3 code. For other programming languages, although the specific commands will be different, the general principles of working with variables will likely be similar.

5.1 READING IN EN4 PROFILES AND PLOTTING THEIR LOCATIONS

```
###  
BSD 3-Clause License  
This code is licensed under the BSD licence.  
British Crown Copyright © 2021, Met Office  
All rights reserved.  
###  
  
import cartopy  
import cartopy.crs as ccrs  
import matplotlib.pyplot as plt  
import netCDF  
from netCDF4 import Dataset  
  
# Load in a month of data:  
  
jan = Dataset('EN.4.2.2.f.profiles.g10.197001.nc')  
  
# Look at all the variables read in:  
  
print(jan)
```

```

# Read in the locations (note, if you just wanted the 10th value you could replace
\:' (which reads everything in) with 10):

lat = jan['LATITUDE'][: ]
lon = jan['LONGITUDE'][: ]

# Plot the locations:

plt.scatter(lon, lat)
plt.show()

# Plot the locations with a map outline:

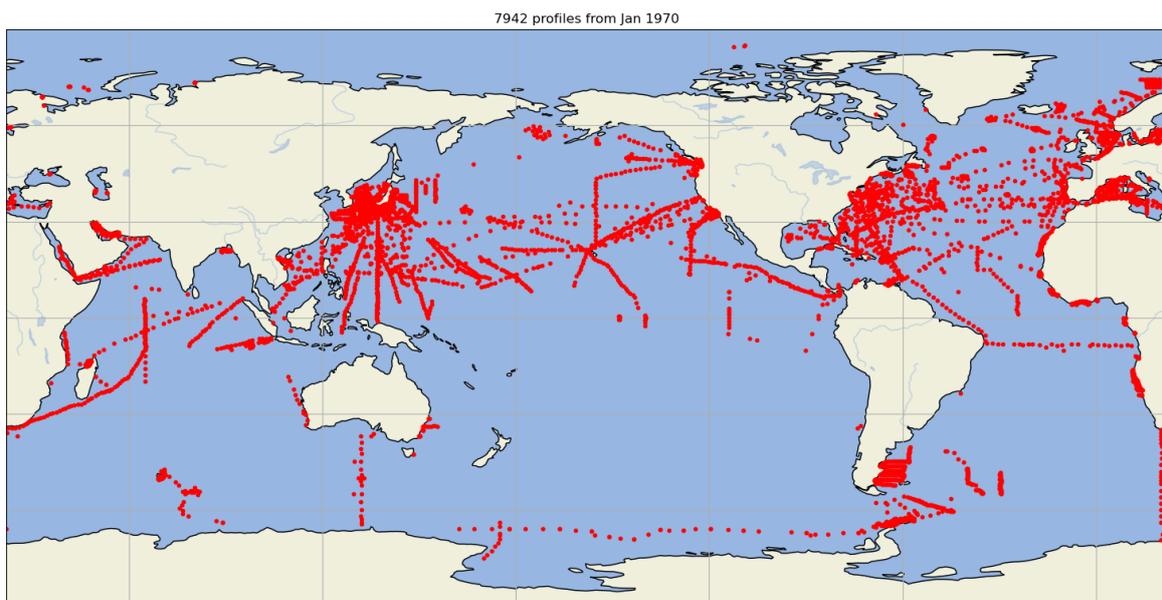
SRC_CRSS = ccrs.PlateCarree()
MAP_CRSS = ccrs.PlateCarree(central_longitude=-158)

fig = plt.figure(figsize=(7, 4))
ax = plt.axes(projection=MAP_CRSS)
ax.set_global()
ax.add_feature(cartopy.feature.LAND)
ax.add_feature(cartopy.feature.OCEAN)
ax.add_feature(cartopy.feature.COASTLINE)
ax.add_feature(cartopy.feature.LAKES, alpha=0.5)
ax.add_feature(cartopy.feature.RIVERS, alpha=0.5)
ax.gridlines()

new_lons, new_lats, _ = MAP_CRSS.transform_points(SRC_CRSS, lons, lats).T
ax.plot(new_lons, new_lats, 'r.')
ax.set_title(f'{jan.dimensions["N_PROF"].size} profiles from Jan 1970')
plt.show()

# Expected output:

```



5.2 MERGING MONTHLY EN4 ANALYSIS FILES INTO AN ANNUAL FILE

```
###
BSD 3-Clause License
This code is licensed under the BSD licence.
British Crown Copyright © 2021, Met Office
All rights reserved.
###

import iris

from iris.util import equalise_attributes, unify_time_units

import os

# Load in the months of data that we are interested in (e.g. the year 2000):
filepath = 'LOCATION_OF_MONTHLY_FILES'
filename = 'EN.4.2.2.f.analysis.109.2000*.nc'
cubes_2000 = iris.load(os.path.join(filepath, filename))

# Ensure the cubes (data from the files) can be combined:
equalise_attributes(cubes_2000)
unify_time_units(cubes_2000)

# Combine the files:
cube_2000 = cubes_2000.concatenate()

# cube_2000 then has a time dimension of 12 and all six variables still separate.
print(cube_2000)

# Expected output (although note the cubes could be listed in a different order):
0: salinity error standard deviation / (1) (time: 12; depth: 42; latitude: 173;
longitude: 360)
1: salinity observation weights / (unknown) (time: 12; depth: 42; latitude: 173;
longitude: 360)
2: sea_water_potential_temperature / (kelvin) (time: 12; depth: 42; latitude: 173;
longitude: 360)
3: sea_water_salinity / (1) (time: 12; depth: 42; latitude: 173;
longitude: 360)
4: temperature error standard deviation / (kelvin) (time: 12; depth: 42; latitude:
173; longitude: 360)
5: temperature observation weights / (unknown) (time: 12; depth: 42; latitude: 173;
longitude: 360)

# To look at the data you can do:
cube_2000.extract('variable_name')[0].data

# Note that this code will not work with profile files which are structured
differently.
```

5.3 CREATING A TIME SERIES OF A SINGLE ANALYSIS LOCATION

```
###
```

[BSD 3-Clause License](#)

This code is licensed under the BSD licence.
British Crown Copyright © 2021, Met Office
All rights reserved.

###

```
import iris
from iris.util import equalise_attributes, unify_time_units
import matplotlib
import matplotlib.pyplot as plt
matplotlib.rcParams.update({'font.size': 12})
import os
import pandas as pd

# Load in the months of data we are interested in (e.g. all data from the 2010s):
filepath = 'LOCATION_OF_MONTHLY_ANALYSIS_FILES'
filename = 'EN.4.2.2.f.analysis.c13.201*.nc'
cubes_2010s = iris.load(os.path.join(filepath, filename))

# Ensure the cubes (data from the files) can be combined:
equalise_attributes(cubes_2010s)
unify_time_units(cubes_2010s)

# Combine the files:
cube_2010 = cubes_2010s.concatenate()

# Extract variables at one point chosen at random (17N and -59E, first depth
level):
pot_temp = cube_2010.extract('sea_water_potential_temperature')[0][:, 0, 100,
300].data
pot_temp_unc = cube_2010.extract('temperature error standard deviation')[0][:, 0,
100, 300].data
temp_obs_inf = cube_2010.extract('temperature_observation_weights')[0][:, 0, 100,
300].data

# Turn this information into time series:
ts_index = pd.date_range("2010-01-01", periods=120, freq="M")
ts = pd.Series(pot_temp - 273.15, index=ts_index)
ts_unc = pd.Series(pot_temp_unc, index=ts_index)
ts_oi = pd.Series(temp_obs_inf, index=ts_index)

# Create some plots (if you want these as individual plots just remove the
plt.subplot() sections and execute each section of plotting code separately):
plt.subplot(2, 2, 1)
ts.plot()
```

```

plt.fill_between(ts_index, (pot_temp - 273.15) + pot_temp_unc,
                 (pot_temp - 273.15) - pot_temp_unc, alpha=0.5)

plt.title('(a)')

plt.ylabel('Potential temperature (degrees C)')

plt.xlabel('Time')

# Note, this plot can be misleading if not adequately sized as it appears that the
# uncertainties are larger at the peaks. The plot directly below it shows that this
# is not the case.

plt.subplot(2, 2, 3)

ts[:12].plot()

plt.fill_between(ts_index[:12], (pot_temp[:12] - 273.15) + pot_temp_unc[:12],
                 (pot_temp[:12] - 273.15) - pot_temp_unc[:12], alpha=0.5)

plt.title('(c)')

plt.ylabel('Potential temperature (degrees C)')

plt.xlabel('Time')

plt.subplot(2, 2, 2)

ts_unc.plot()

plt.title('(b)')

plt.ylabel('Analysis error standard deviation (degrees C)')

plt.xlabel('Time')

plt.subplot(2, 2, 4)

ts_oi.plot()

plt.title('(d)')

plt.ylabel('Temperature observation weights')

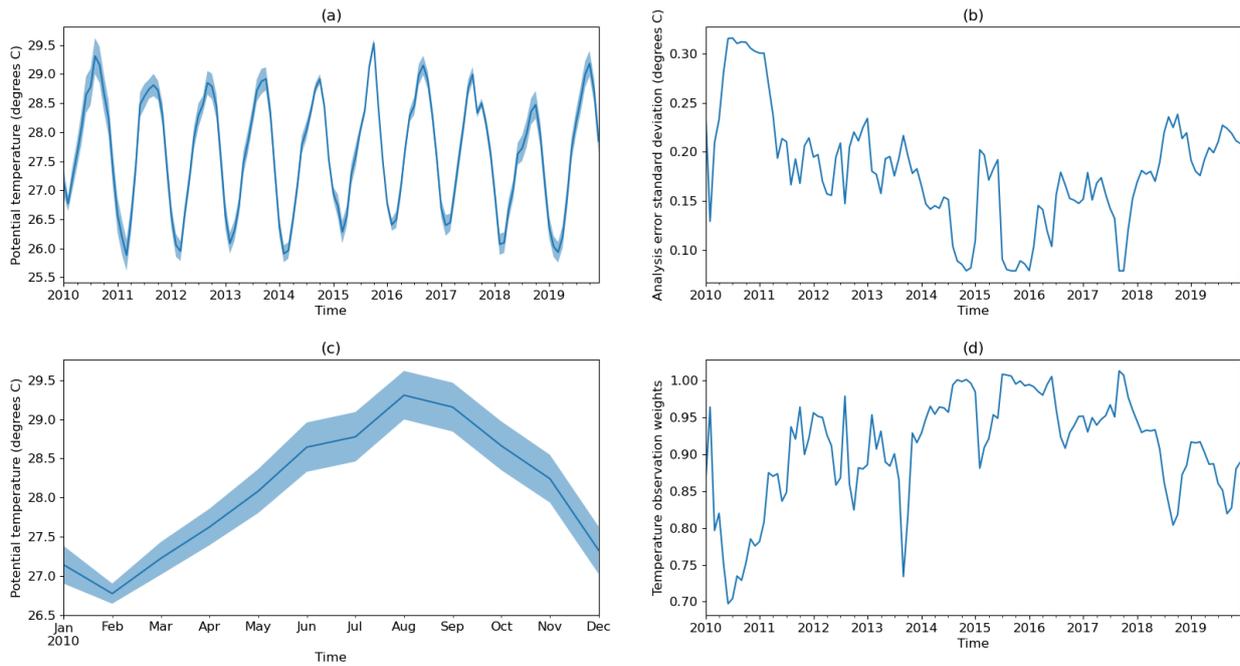
plt.xlabel('Time')

plt.show()

# Because of the link between analysis error standard deviations and the presence
# of observations, the two are approximately inversely related, that is, the standard
# deviations are largest when the observation weights are smallest.

# Expected output:

```



Note that this is a single analysis point and therefore we can take uncertainties as they are provided. Analysis uncertainties cannot be combined assuming independence as this will result in a combined uncertainty that is too small. The uncertainties do not currently have a defined correlation structure in order to do a more sophisticated combined uncertainty calculation, but this is something we hope to provide in a later version of EN4.

5.4 PLOTTING AN ANALYSIS FIELD AND THE OBSERVATION WEIGHTS

###

[BSD 3-Clause License](#)

This code is licensed under the BSD licence.

British Crown Copyright © 2021, Met Office

All rights reserved.

###

```
import cartopy
import cartopy.crs as ccrs
import iris
import matplotlib
import matplotlib.pyplot as plt
matplotlib.rcParams.update({'font.size': 15})
import os

# Load the file and extract the variables of interest:
file_path = 'LOCATION_OF_ANALYSIS_FILE'
analysis = iris.load(os.path.join(file_path, 'EN.4.2.2.f.analysis.c14.202104.nc'))

temps = analysis.extract('sea_water_potential_temperature')[0]
obsweights = analysis.extract('temperature observation weights')[0]

# Get the boundaries of the coordinates (e.g., the area which a grid box covers):
temps.coord('latitude').guess_bounds()
```

```

temps.coord('longitude').guess_bounds()
obsweights.coord('latitude').guess_bounds()
obsweights.coord('longitude').guess_bounds()

# Turn the iris cube data structures into arrays
gridlons = obsweights.coord('longitude').contiguous_bounds()
gridlats = obsweights.coord('latitude').contiguous_bounds()

# Extract a single analysis level (level 23 is 596 - 725m):
obstemps = temps.data[0, 23, :, :] - 273.15
weights = obsweights.data[0, 23, :, :]

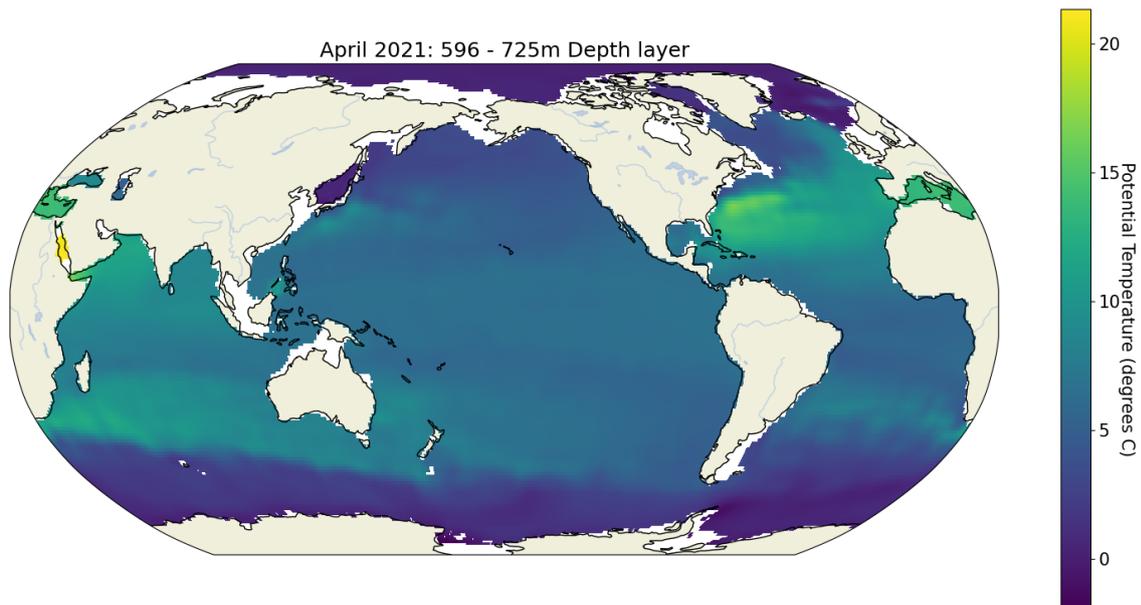
# Create the figures:
# Temperatures:
fig = plt.figure(figsize=(7, 4))
ax = plt.axes(projection=ccrs.Robinson(central_longitude = -158))

# Define the coordinate system that the grid longitudes and grid latitudes are on:
tran = ccrs.PlateCarree()
plt.pcolormesh(gridlons, gridlats, obstemps, transform = tran)
cbar = plt.colorbar()
cbar.set_label('Potential Temperature (degrees C)', rotation=270,labelpad=12)

ax.add_feature(cartopy.feature.LAND)
ax.add_feature(cartopy.feature.COASTLINE)
ax.add_feature(cartopy.feature.LAKES, alpha=0.5)
ax.add_feature(cartopy.feature.RIVERS, alpha=0.5)
ax.set_title('April 2021: 596 - 725m Depth layer')
plt.show()

# Expected output:

```

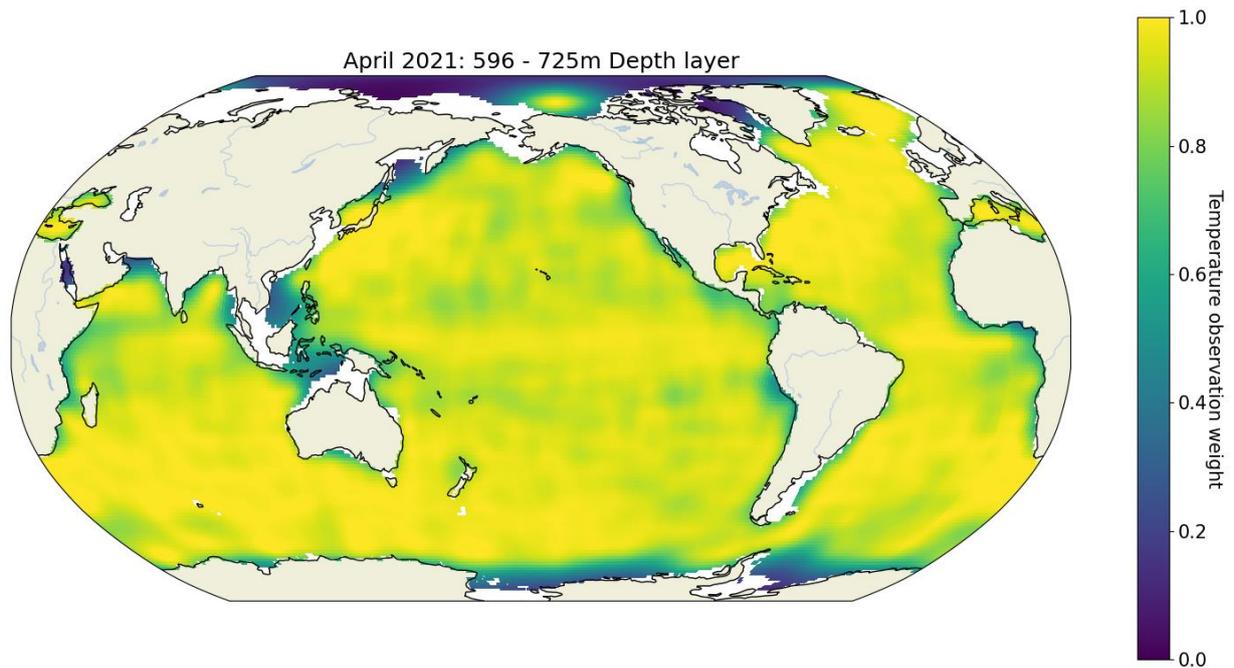


```
# Temperature observation weights:
fig = plt.figure(figsize=(7, 4))
ax = plt.axes(projection=ccrs.Robinson(central_longitude = -158))

plt.pcolormesh(gridlons, gridlats, weights, transform = tran)
cbar = plt.colorbar()
plt.clim(0,1)
cbar.set_label('Temperature observation weight', rotation=270,labelpad=12)

ax.add_feature(cartopy.feature.LAND)
ax.add_feature(cartopy.feature.COASTLINE)
ax.add_feature(cartopy.feature.LAKES, alpha=0.5)
ax.add_feature(cartopy.feature.RIVERS, alpha=0.5)
ax.set_title('April 2021: 596 - 725m Depth layer')
plt.show()

# Expected output:
```



5.5 PLOTTING A SALINITY PROFILE

```

###
BSD 3-Clause License
This code is licensed under the BSD licence.
British Crown Copyright © 2021, Met Office
All rights reserved.
###

import iris
import matplotlib
import matplotlib.pyplot as plt
matplotlib.rcParams.update({'font.size': 15})
import os

# Load the file:

file_path = 'LOCATION_OF_PROFILE_FILE'

file_name = 'EN.4.2.2.f.profiles.109.202008.nc'

cubes = iris.load(os.path.join(file_path, file_name))

# Extract the variables of interest:

depth = cubes.extract('corrected depth')[0]

psal = cubes.extract('corrected practical salinity')[0]

# Plot a profile (chosen at random):

fig1 = plt.figure()

ax1 = fig1.add_subplot(1,1,1)

ax1.plot(psal[29].data, depth[29].data) # 29 is the randomly chosen profile number

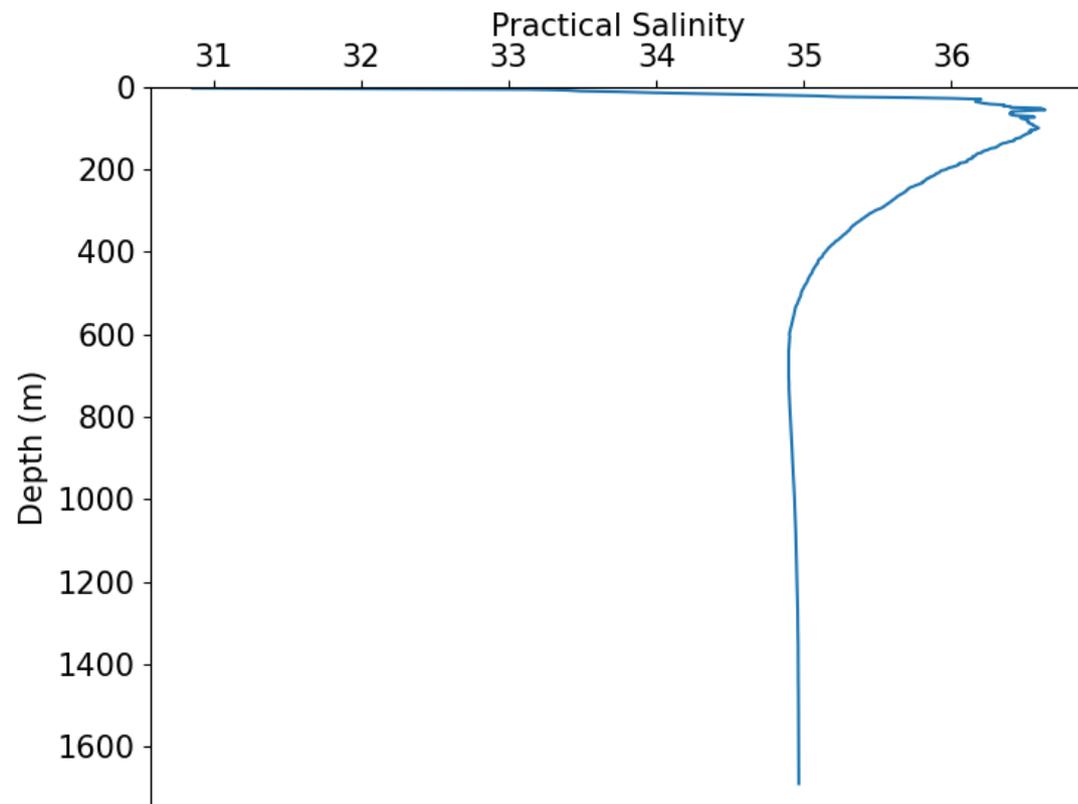
ax1.set_ylabel('Depth (m)')

```

```

ax1.set_ylim((0, 1750)[::-1]) # This reverses the y-axis (i.e. deep at the bottom)
ax1.set_xlabel('Practical Salinity')
ax1.xaxis.set_label_position('top')
ax1.xaxis.set_ticks_position('top')
plt.show()
# Expected output:

```



5.6 SELECTING ONLY PROFILES THAT PASS QC

```

###
BSD 3-Clause License
This code is licensed under the BSD license.
British Crown Copyright © 2021, Met Office
All rights reserved.
###

import cartopy
import cartopy.crs as ccrs
import matplotlib
import matplotlib.pyplot as plt
matplotlib.rcParams.update({'font.size': 15})
from netCDF4 import Dataset
import numpy as np
import os

file_path = 'LOCATION_OF_PROFILE_FILE'

file_name = 'EN.4.2.2.f.profiles.c13.199805.nc'

profs = Dataset(os.path.join(file_path, file_name))

```

```

# Extract the QC flag for the variable of interest:
potm_qc = profs.variables['PROFILE_POTM_QC'][:].astype('unicode')

# Find only the locations where the QC flag is good (1):
good_profs = np.nonzero(potm_qc == '1')[0]

# To select only these locations:
lats = profs['LATITUDE'][good_profs]
lons = profs['LONGITUDE'][good_profs]

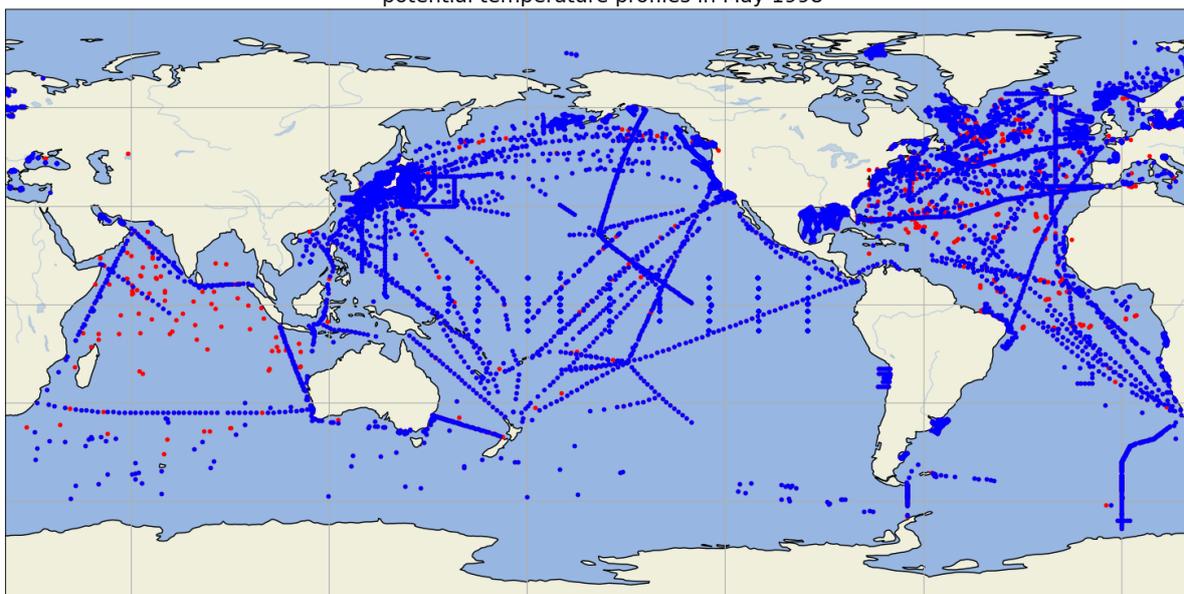
# To select all locations:
lats_all = profs['LATITUDE'][:]
lons_all = profs['LONGITUDE'][:]

# Plot the locations with a map outline (plotting all locations first and then
overplotting the good locations allows the bad locations to still show):
SRC_CRSS = ccrs.PlateCarree()
MAP_CRSS = ccrs.PlateCarree(central_longitude=-158)
fig = plt.figure(figsize=(7, 4))
ax = plt.axes(projection=MAP_CRSS)
ax.set_global()
ax.add_feature(cartopy.feature.LAND)
ax.add_feature(cartopy.feature.OCEAN)
ax.add_feature(cartopy.feature.COASTLINE)
ax.add_feature(cartopy.feature.LAKES, alpha=0.5)
ax.add_feature(cartopy.feature.RIVERS, alpha=0.5)
ax.gridlines()
new_lons_all, new_lats_all, _ = MAP_CRSS.transform_points(SRC_CRSS, lons_all,
lats_all).T
new_lons, new_lats, _ = MAP_CRSS.transform_points(SRC_CRSS, lons, lats).T
ax.plot(new_lons_all, new_lats_all, 'r.')
ax.plot(new_lons, new_lats, 'b.')
ax.set_title(f'Locations of good (blue) and bad (red) \n potential temperature
profiles in May 1998')
plt.show()

# Expected output:

```

Locations of good (blue) and bad (red)
potential temperature profiles in May 1998



5.7 PLOTTING PROFILES WITHIN A GRID BOX AND FLAGGING LEVELS THAT HAVE FAILED QC

```
###  
BSD 3-Clause License  
This code is licensed under the BSD licence.  
British Crown Copyright © 2021, Met Office  
All rights reserved.  
###  
  
import matplotlib  
import matplotlib.pyplot as plt  
matplotlib.rcParams.update({'font.size': 15})  
from netCDF4 import Dataset  
import numpy as np  
import os  
  
# Load in the files:  
  
file_path = 'LOCATION_OF_FILE'  
  
file_name_p = 'EN.4.2.2.f.profiles.c14.201901.nc'  
  
profs = Dataset(os.path.join(file_path, file_name_p))  
  
# Choose a grid box to look for profiles in (although note that an analysis grid  
box will not only be influenced by the profiles that fall in that exact location,  
profiles from surrounding grid boxes will also affect the value):  
  
lats = profs.variables['LATITUDE'][:]  
lons = profs.variables['LONGITUDE'][:]  
prof_locs = np.where(np.logical_and(np.logical_and(lats >= -28.5, lats <= -27.5),  
                                   np.logical_and(lons >= 39.5, lons <= 40.5)))  
  
# Now get the temperatures and depths of these profiles and their QC flags:
```

```

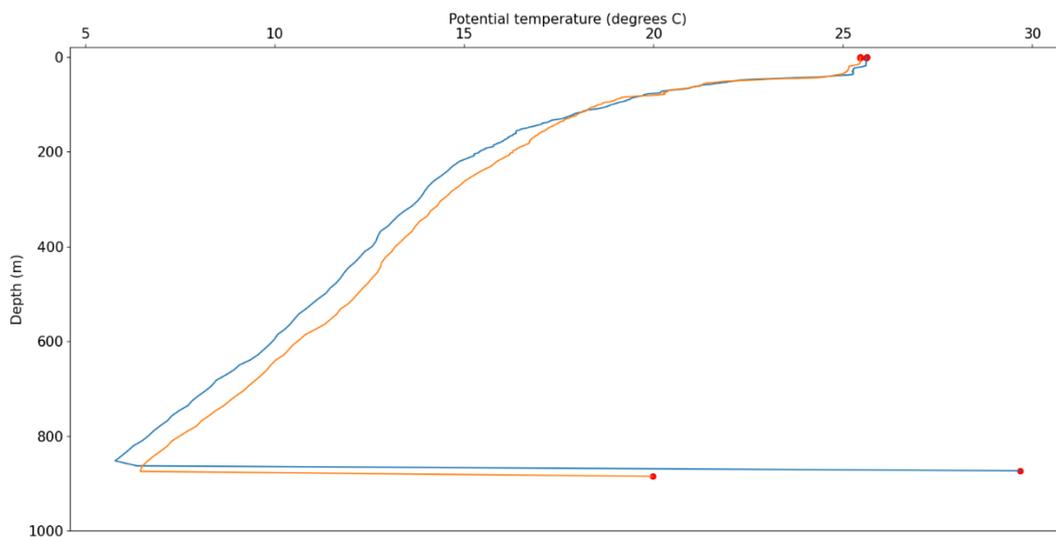
temps = profs['POTM_CORRECTED'][prof_locs]
depths = profs['DEPH_CORRECTED'][prof_locs]
temp_level_qc = profs['POTM_CORRECTED_QC'][prof_locs].astype('unicode')

# Now plot these profiles:
for prof in range(len(prof_locs[0])):
    plt.plot(temps[prof], -depths[prof])
    bad_locs = np.where(temp_level_qc[prof] == '4')
    plt.scatter(temps[prof][bad_locs], -depths[prof][bad_locs], c='red')

plt.xlabel('Temperature (degrees C)')
plt.ylabel('Depth (m)')
plt.show()

# Expected output:

```



5.8 IDENTIFYING WHICH QC CHECKS HAVE BEEN FAILED

```
###
```

[BSD 3-Clause License](#)

This code is licensed under the BSD licence.
British Crown Copyright © 2021, Met Office
All rights reserved.

```
###
```

```

import os
from netCDF4 import Dataset
import numpy as np
import random

```

```
# Load in the file and extract the variables of interest:
```

```
file_path = 'LOCATION_OF_PROFILE_FILES'
```

```
file_name = 'EN.4.2.2.f.profiles.c13.200803.nc'
```

```

profs = Dataset(os.path.join(file_path, file_name))

qc_profs = profs['QC_FLAGS_PROFILES'][:]
qc_levs = profs['QC_FLAGS_LEVELS'][:]

# Find some occasions where qc_profs is nonzero (meaning at least one QC check has
been failed):
bad_qc_prof_locs = np.where(qc_profs != 0)[0]

# Choose one of these profiles at random:
random.choice(bad_qc_prof_locs) # 19406

# Look at which bits are set:
qc_profs[19406] # 2097154

# Make this binary (and reverse the order so that bit zero is on the left hand
side)
bin(2097154)[::-1]

# Expected output is 010000000000000000000001b0 - this shows bits 1 and 21 are set
(Python counts from zero) and this tells us the salinity profile was rejected
because over half of its levels were rejected.

# Can use a similar approach for qc_levs (excluding values of 33562627 means we
don't select the values that usually indicate missing data):
bad_qc_lev_locs = np.where(np.logical_and(qc_levs != 0, qc_levs != 33562627))
np.shape(bad_qc_lev_locs) # (2, 158185)
random.randrange(158185) # 128530

bad_qc_lev_locs[0][128530] # 37724 - this is the profile number
bad_qc_lev_locs[1][128530] # 0 - this is the level of the profile
bin(qc_levs[37724][0])[::-1]

# Expected output is 100000001b0 - this shows bits 0 and 8 are set meaning the
temperature level should be rejected because it's a bathythermograph measurement
shallower than 4m.

```

5.9 WORKING WITH THE INST_REFERENCE AND PROJECT_NAME VARIABLES

5.9.1 Finding delayed mode Argo profiles

```

###
BSD 3-Clause License
This code is licensed under the BSD licence.
British Crown Copyright © 2021, Met Office
All rights reserved.

```

```

###

import netCDF4
from netCDF4 import Dataset
import numpy as np
import os

# Load in an EN4 file from approximately 6 months before the source data were
downloaded, then the expectation would be that there is a mixture of real time,
real time adjusted and delayed mode Argo profiles:

file_path = 'LOCATION_OF_PROFILE_FILE'

file_name = 'EN.4.2.2.f.profiles.g10.202010.nc'

profs = Dataset(os.path.join(file_path, file_name))

# Extract the project name variable:

proj_name = netCDF4.chartostring(profs['PROJECT_NAME'][:].astype('|S1'))

# Create an array containing the first five characters of this variable (by writing
a function to extract the first five characters of the project name and then
applying this over the whole variable):

def extract_proj(proj):

    return proj[:5]

vfunc_proj = np.vectorize(extract_proj)

pn_short = vfunc_proj(proj_name)

# Now find the locations (and counts) of different modes of Argo profiles

adj = np.where(pn_short == 'ARGOA')

print(np.shape(adj)) # Expected output is (1, 6309)

delayed_mode = np.where(pn_short == 'ARGOD')

print(np.shape(delayed_mode)) # Expected output is (1, 1732)

real_time = np.where(pn_short == 'ARGOR')

print(np.shape(real_time)) # Expected output is (1, 5625)

```

5.9.2 Finding all XBT profiles and their types

```

###
BSD 3-Clause License
This code is licensed under the BSD licence.
British Crown Copyright © 2021, Met Office
All rights reserved.
###
from collections import Counter
import netCDF4

```

```

from netCDF4 import Dataset
import numpy as np
import os

# Load in an EN4 file and extract the variables of interest:
file_path = 'LOCATION_OF_PROFILE_FILES'
file_name = 'EN.4.2.2.f.profiles.109.199507.nc'

profs = Dataset(os.path.join(file_path, file_name))

proj_name = netCDF4.chartostring(profs['PROJECT_NAME'][:].astype('|S1'))
inst_ref = netCDF4.chartostring(profs['INST_REFERENCE'][:].astype('|S1'))

# Only WOD and GTSPP profiles can contain XBT data - use similar functions to
example 9.1 to extract the WOD and GTSPP profiles:

def extract_proj(proj):
    return proj[:8]

vfunc_proj = np.vectorize(extract_proj)
pn_short = vfunc_proj(proj_name)

# Now extract the XBT profiles:
wod_xbt_profs = np.where(pn_short == 'WOD09XBT')[0]
gtspp_xbt_profs = np.where(pn_short == 'GTSPPXB ')[0]

# Now use the information in INST_REFERENCE to determine the type of XBTs:

# For GTSPP this is simple as the instrument type is just the value contained in
# the INST_REFERENCE field:
gtspp_xbt_types = (inst_ref[gtspp_xbt_profs]).astype(int)
Counter(gtspp_xbt_types)

# Expected output: Counter({52: 89, 41: 5, 1: 4, 0: 1}) - using WMO code table 1770
shows that there are 89 Sippican Deep Blue XBTs, 5 Sippican T7s, 4 Sippican T4s and
one instrument of unknown type.

# For WOD, another sub-selection function needs to be run:
def extract_wod_inst(inst):
    return inst[-4:-1]

```

```

vfunc_inst = np.vectorize(extract_wod_inst)
wod_xbt_types = vfunc_inst(inst_ref[wod_xbt_profs])
Counter(wod_xbt_types)

# Expected output: Counter({'002': 3416, '207': 1740, '214': 360, '216': 189,
'215': 153, '208': 116, '209': 80, '210': 73, '219': 31, '236': 22, '211': 10,
'235': 2, '226': 2, '212': 2, '201': 1})

# For this we need WOD table v5
(https://www.ncei.noaa.gov/data/oceans/woa/WOD/CODES/PDF/v_5_instrument.pdf) and
this tells us there are 3416 XBTs of unknown type, 1740 Sippican T7s, 360 Sippican
Deep Blues, 189 TSK T6s, 153 TSK T4s, 116 Sippican T4s and so on.

```

6 FAQs

6.1 HOW DOES EN4 COMPARE WITH OTHER DATASETS?

EN4 is a monthly updating, quality controlled, profile and analysis dataset, from 1900 to the present day. The analysis files use an optimal interpolation algorithm, meaning that they are not model based, but are instead based on the information available in the profiles and the background fields. In the persistent absence of observations, these background fields relax back to climatology. The quality control checks are automatic, although any problem profiles identified by developers or users are added to a reject-list to ensure they are removed in future versions of the dataset.

For a comparison of EN4 and other data sets, see the [HadIOD.1.2.0.0 Product User Guide](#).

6.2 WHAT DO THE FILE NAMES MEAN?

EN4 files are named using the following convention:

dataset.X.Y.Z.θ.type.E.date.extension

For EN4.2.2 these elements are as follows:

dataset	EN
X	The major dataset number, 4.
Y	Version number that is incremented when there is a major change to the methodology used to construct the dataset, 2.
Z	Version number that is incremented when there is a minor change to the dataset, 2.
θ	Set to 'p' if the file is a preliminary version or 'f' if the data file has been finalised.
type	This is 'profiles' if the file contains profile data. For objective analysis data it is set to 'analysis' (data files) or 'analyses' (zip files).
E	This indicates the bias corrections applied to the data, one of l09, g10, c13 or c14.
date	The date in the form YYYYMM (YYYY = year, MM = month). For the yearly zip files the month is not included.
extension	Indicates the file type - this is 'nc' (NetCDF) for all EN4 data files and 'zip' for zip files.

6.3 HOW DOES EN.4.2.2 DIFFER FROM ITS PREDECESSOR?

EN.4.2.2 differs from EN.4.2.1 in the following ways:

- A fresh download and processing of the source data. EN.4.2.1 data was predominantly downloaded in 2016 and 2017. EN.4.2.2 data were predominantly downloaded in October 2020 with a fresh download of 2019 – 2021 data from Argo and GTSP in March 2021. For monthly updates the most up to date data are downloaded from Argo and GTSP for the current and two previous months.
- The updating of the values used for the Levitus et al (2009) corrections. The values used are those given in the 'July 2019' Levitus corrections links of <https://www.ncei.noaa.gov/mechanical-bathythermograph-mbt-corrections> and <https://www.ncei.noaa.gov/expendable-bathythermograph-xbt-corrections>.
- The use of new Gouretski and Cheng (2020) MBT corrections in the .g10. and .c14. ensemble members. In .g10. these replace the Gouretski and Reseghetti (2010) MBT corrections.
- The addition of corrections to T10 XBTs in the .g10. corrections.
- The removal of the temperature dependent part of the depth correction in the .g10. corrections.
- There is one variable change in the .g10. data in EN.4.2.2 compared to EN.4.2.1. In EN.4.2.1 the .g10. data contained a variable 'stretch_corrections', which provided the values that the uncorrected depths had been multiplied by in order to get the corrected depths. In EN.4.2.2 the .g10. data contain the variable 'depth_corrections' instead, this provides the value that has been subtracted off the uncorrected depths in order to get the corrected depths which are those included in the files.
- The addition of Cheng et al (2014) XBT bias corrected data (paired with Gouretski and Cheng (2020) MBT corrections) in the .c14. ensemble member.
- The addition of Cowley et al (2013) XBT bias corrected data (paired with the Levitus et al (2009) MBT corrections) in the .c13. ensemble member.
- The addition of licence and reference attributes to all netCDF files to aid correct citations and use.
- The addition to the EN4 reject list of salinity data from eight TAO buoys between 2015 and 2016 after spurious salinity values were found by a user in EN.4.2.1.
- The addition to the EN4 reject list of a single early Argo float after stuck depths were reported by a user in EN.4.2.1.
- The removal of data marked as being from MBTs prior to their development in 1938.
- The removal of data marked as being from Argo in 1956.
- The removal of a re-classification of T7 probes going deeper than 840m to be T5s from 1990 onwards following communication with other data set developers on T7 maximum depths.
- Additional quality assurance checks were added in during this data processing – these included more checks of the analyses' outputs:
 - Comparison to the minimum and maximum values seen previously in the data set (with a tolerance of 3).
 - Comparison of like for like fields to the previous version of the data set (with a tolerance of 8). Artefacts flagged by this check were assessed and deemed to be correct in EN.4.2.2, mainly caused by additional data that were not present in EN.4.2.1.

6.4 HOW CAN I ACCESS OLDER VERSIONS OF THE EN DATASET?

We strongly encourage the use of the most up to date version of EN4 in any research. However, if you require older versions of the dataset for comparison or replication of previous work please [contact us](#).

6.5 CAN I USE EN4 FOR TREND ANALYSIS?

EN4 is an observation based data set and therefore coverage varies in both time and space. For the complete fields provided in the analyses we strongly encourage users to look at the 'observation weights' variables. These will inform users how much the analysis value has been influenced by observations (values closer to one) and how much it has been determined by background fields (values closer to zero). We would not encourage the use of EN4 analyses for trend analysis in areas where the observation weights are low.

6.6 IS X SPECIFIC DATA SOURCE IN EN4?

EN4 takes data from ASBO, Argo, GTSP and WOD18. These sources themselves draw from other sources; it is therefore recommended that you check the information on these four input sources to determine whether EN4 contains the data you are interested in.

6.7 HOW DO YOU DEAL WITH POSSIBLE DUPLICATES?

The process for duplicate removal is described in section 2.2 of [Good et al., 2013](#) and is built from the work of Grønnell and Wijffels, 2008. Briefly, the steps look to remove duplicates that arise from:

1. Argo profiles that are also in WOD and GTSP – prefer the profile from Argo;
2. Profiles that have the exact (or almost exact) time and location;
3. Profiles that have exactly the same depths, temperatures and salinities;
4. Profiles that are very similar, but not exactly the same (only profiles with 5 or more levels and near to each other (within 1 degree latitude and longitude and 1 day in time) are checked).

6.8 DO YOU INCLUDE QUALITY CONTROL FLAGS FROM YOUR INPUT SOURCES?

Predominantly EN4 discards the QC flags of its input sources and performs its own automatic quality control for consistency. Table 1 of Good et al., 2013 gives an overview of these checks. The exception to this is that Argo QC flags are retained, and data flagged as bad or probably bad by Argo will also be flagged in EN4. We also take into account the Argo grey lists and comparisons with altimetry data as well as maintaining a manual reject list file of profiles identified to be bad by data set maintainers and users.

6.9 HOW DO YOU DEAL WITH ARGO DATA WITH A SALINITY DRIFT?

Certain batches of Argo floats have been found to experience a salinity drift. Many of these are flagged for rejection in the Argo grey list files and are therefore flagged as bad accordingly in EN4. EN4 also uses delayed mode Argo data where possible in which these data will have been either corrected or flagged for rejection. Some floats will not have been identified at the time of the creation of EN4.2.2 and therefore will still exhibit this salinity drift. If you identify any such floats, please notify us.

6.10 WHICH INSTRUMENTS ARE USED IN EN4?

EN4 takes data from all instruments used in World Ocean Database, apart from those known to only be taking measurements at the ocean surface. This means EN4 will include profiles from Argo and other profiling floats, bathythermographs, bottles, CTDs, drifting buoys, gliders, moored buoys, STDs, instruments attached to animals, undulating oceanographic recorders and XCTDs. For an overview of these and other instruments we recommend the [World Ocean Database 2018 Introduction](#).

6.11 WHICH INSTRUMENTS ARE BIAS CORRECTED IN EN4?

The only instruments we apply bias corrections to in EN4 are MBTs and XBTs, although some data processing will have been undertaken before the data reach us e.g. Argo float delayed mode data may have been adjusted.

6.12 WHICH SET OF BIAS CORRECTIONS SHOULD I USE FOR MY APPLICATION?

EN4 provides four ensemble members, l09, g10, c13 and c14, each with a different combination of MBT and XBT bias corrections (see [section 4](#)). We recommend that you look at all four ensemble members to see what impact this has on your conclusions.

6.13 WHAT IS THE REFERENCE PRESSURE FOR POTENTIAL TEMPERATURE?

0 dbar.

6.14 WHAT ARE THE UNITS OF THE VARIABLES IN EN4?

In the EN4 profile files all temperatures are given in degrees Celsius, salinities are unitless and depths are given in metres. In the analysis files temperatures are recorded in Kelvin, but with a 273.15 offset built into the netCDF files, this offset won't always need to be subtracted by the user, it is dependent on the software you use to read netCDF files in. The salinities are unitless and the depths are given in metres.

6.15 HOW IS POTENTIAL TEMPERATURE CALCULATED WHEN NO SALINITY IS PRESENT?

Not all temperature measurements will have accompanying salinity measurements. Where this is the case, we use a salinity value from the background fields. If no background salinity value is available, then we use a salinity value of 35.0 but flag the potential temperature value as untrustworthy.

6.16 HOW DO I FIND THE DEPTH LEVELS OF THE EN4 ANALYSES?

The EN4 analysis depth levels are given [here](#). If you want to view these using a Linux interface you can make use of the NCO tool ncdump. Typing `ncdump -v depth NAME_OF_FILE.nc` will give you the depths and `ncdump -v depth_bnds NAME_OF_FILE.nc` will give you the bounds of these levels. If you are a Windows user then the method of extracting this information will be dependent on your choice of software, but depth and depth_bnds will still be the variables of interest.

6.17 HOW MANY LEVELS ARE IN THE EN4 PROFILE FILES?

EN.4.2.2 has a maximum of 400 levels in a profile file. If source data profiles have more levels than this then they are thinned until there are 400 levels or less. This thinning begins at the deepest

depths, thinning to observations approximately every 50m below 2000m. Once approximately 400 levels remain, thinning stops.

6.18 WHAT DATA CONVENTIONS DOES EN4 FOLLOW?

EN.4.2.2 data are served as netCDF files which are broadly compliant with the CF conventions standard 1.8. However, there are variables which are non-compliant, notably owing to a lack of appropriate long_names or units. This is something we will seek to resolve in future versions of the EN4 data set.

6.19 WHAT ARE THE TERMS AND CONDITIONS OF USING EN4?

EN4 is available for private and research use and is served under a [non-commercial government licence](#). You can find the terms and conditions for using EN4 here https://www.metoffice.gov.uk/hadobs/en4/terms_and_conditions.html.

6.20 WHY DO THE ANALYSIS FILES START AT -83N?

To save space in the analysis files that we serve we remove the most southerly 7 degrees latitude owing to there being no ocean in this zone.

6.21 ARE THERE UNCERTAINTIES IN EN4?

EN4 analysis files do incorporate uncertainty variables for temperatures and salinities. These uncertainties are the analysis error standard deviations of the analysis fields. The uncertainties are not independent of each other, but also do not have a defined correlation structure to allow the combining of uncertainties. This is something we aim to provide in a future version of the EN dataset.

EN4 profile files do not incorporate uncertainties, however, values of random uncertainties for each instrument type can be obtained from both tables 2 and 3 in the [HadIOD.1.2.0.0 user guide](#) and tables 1 and 2 of the International Quality controlled Ocean Database (IQuOD) [uncertainties paper](#).

Some element of structural uncertainty is represented in EN4 through the provision of four ensemble members, using different XBT and MBT bias correction schemes, see [section 4](#) for further information on these.

6.22 WHAT IS THE PRECISION OF THE DATA IN EN4?

EN4 data are provided to a greater number of decimal places than precision would be expected to extend to, owing to the storage mechanism in netCDF files where all numbers are stored as floats. The tables linked in [FAQ 6.20](#) give an idea of the expected uncertainties in values measured by different instruments, we would therefore not recommend assuming a precision to a greater number of decimal places than these uncertainties are given for.

6.23 WHERE CAN I FIND MORE INFORMATION ABOUT EN4?

For further information on EN4 we recommend reading the paper [Good et al., 2013](#).

6.24 HOW CAN I GIVE FEEDBACK?

If you're using EN4 we would be very grateful for your feedback, whether that is on aspects you particularly like or things you would like to see added or done differently. Please [contact us](#) with any feedback. Likewise, if you find data that you believe should be flagged as bad, or data that are flagged that you believe shouldn't be then please get in touch.

7 GLOSSARY OF ACRONYMS AND SHORTHAND

C13 = Ensemble member with Cowley et al., 2013 XBT corrections and Levitus et al., 2009 MBT corrections.

C14 = Ensemble member with Cheng et al., 2014 XBT corrections and Gouretski and Cheng, 2020 MBT corrections.

G10 = Ensemble member with Gouretski and Reseghetti, 2010 XBT corrections and Gouretski and Cheng, 2020 MBT corrections.

GTSP = Global Temperature and Salinity Profile Programme.

HadiOD = Hadley Centre Integrated Ocean Database.

IQuOD = International Quality controlled Ocean Database.

L09 = Ensemble member with Levitus et al., 2009 XBT corrections and Levitus et al., 2009 MBT corrections.

MBT = Mechanical BathyThermograph.

QC = Quality Control.

STD = Salinity, Temperature, Depth instrument.

WOD = World Ocean Database.

XBT = eXpendable BathyThermograph.

8 REFERENCES AND ACKNOWLEDGEMENTS

The writing of this user guide benefited greatly from user guides from other, similar products, notably the HadiOD.1.2.0.0 user guide and the Argo User's Manual, version 3.4. The latter was used heavily in the creation of Table 1.

The EN.4.2.2 data set was created using Argo, ASBO, GTSP and WOD18 data. The Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (<https://argo.ucsd.edu>, <https://www.ocean-ops.org>). The Argo Program is part of the Global Ocean Observing System.

The EN.4.2.2 data set makes use of the results of quality control procedures that use altimeter data developed by Stephanie Guinehut. The quality control results were obtained from <https://data-argo.ifremer.fr/etc/argo-ast9-item13-AltimeterComparison/>. The technique is described in Guinehut et al. (2009).

Argo Data Management Team (2021). **Argo user's manual V3.4**. <https://doi.org/10.13155/29825>.

Argo (2000): Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. <https://doi.org/10.17882/42182>.

ASBO sources:

- Hydrobase - Curry (2001).
- Beaufort gyre experiment - The data were collected and made available by the Beaufort Gyre Exploration Program based at the Woods Hole Oceanographic Institution (<https://www2.whoi.edu/site/beaufortgyre/>).
- North Pole Environmental Observatory (NPEO; <http://psc.apl.washington.edu/northpole>) - NSF Grants OPP-9910305, OPP-0352754, and ARC-0634226.
- Freshwater switchyard of the Arctic project - M. Steele, W. Smethie, and P. Schlosser, CTD data from the Arctic Switchyard Project, Office of Polar Programs, National Science Foundation: <http://psc.apl.washington.edu/switchyard/>.
- Nansen and Amundsen Basins Observational System (NABOS), Canadian Basin Observational System (CABOS) (<http://nabos.iarc.uaf.edu/>).
- Physical and Chemical Properties from Selected Expeditions in the Arctic Ocean (Swift 2002; [EOL data archive -- Physical and Chemical Properties from Selected Expeditions in the Arctic Ocean \(ucar.edu\)](http://eol.jhu.edu/data/physical-chemical-properties)).

Cheng, L, J. Zhu, R. Cowley, T. Boyer, and S. Wijffels, 2014: Time, Probe Type and Temperature Variable Bias Corrections to Historical Expendable Bathythermograph Observations. *Journal of Atmospheric and Oceanic Technology*. 31. 1793 – 1825. Doi: 10.1175/JTECH-D-13-00197.1.

Cowley, R, S. Wijffels, L. Cheng, T. Boyer, and S. Kizu, 2013: Biases in Expendable Bathythermograph Data: A new view based on historical side-by-side comparisons. *Journal of Atmospheric and Oceanic Technology*. 30. 1195 – 1225. Doi: 10.1175/JTECH-D-12-00127.1

Curry, R., 2001, HydroBase 2 – A database of hydrographic profiles and tools for climatological analysis. Technical reference.

Gouretski, Victor., and Franco Reseghetti, 2010: On depth and temperature biases in bathythermograph data: development of a new correction scheme based on analysis of a global ocean database. *Deep-sea Research*, 57, 6, doi: <http://dx.doi.org/10.1016/j.dsr.2010.03.11>

Gouretski, Victor., and Lijing Cheng, 2020: Correction for Systematic Errors in the Global Dataset of Temperature Profiles from Mechanical Bathythermographs. *Journal of Atmospheric and Oceanic Technology*, 37, 841 – 855, doi: 10.1175/JTECH-D-19-0205.1

Gronell, A., and S. E. Wijffels (2008), A semiautomated approach for quality controlling large historical ocean temperature archives, *J. Atmos. Oceanic Technol.*, 25, 990–1003, doi:10.1175/JTECHO539.1.

GTSP: Sun, C. and Co-Authors (2010). "The Data Management System for the Global Temperature and Salinity Profile Programme" in *Proceedings of OceanObs.09: Sustained Ocean Observations and Information for Society (Vol. 2)*, Venice, Italy, 21-25 September 2009, Hall, J., Harrison, D.E., & Stammer, D., Eds., ESA Publication WPP-306, [doi:10.5270/OceanObs09.cwp.86](https://doi.org/10.5270/OceanObs09.cwp.86)

Guinehut, S., C. Coatanoan, A.-L. Dhoms, P.-Y. Le Traon and G. Larnicol, 2009. On the use of satellite altimeter data in Argo quality control, *Journal of Atmospheric and Oceanic Technology*, 26, 395-402, DOI: 10.1175/2008JTECHO648.1

Hanawa et al., 1995. A new depth-time equation for Sippican or TSK T-7, T-6 and T-4 expendable bathythermographs (XBT). *Deep-Sea Res.* 42, 1423–1451.

Kizu et al., 2005: Inter-manufacturer difference and temperature dependency of the fall-rate of T-5 expendable bathythermograph. *Journal of Oceanography.* 61 (5), 905–912.

Levitus et al., 2009: Global Ocean Heat Content 1955-2008 in light of recently revealed instrumentation problems. *Geophysical Research Letters*, 36, L07608, doi: <http://dx.doi.org/10.1029/2008GL037155>

WOD18: Boyer, T.P., O.K. Baranova, C. Coleman, H.E. Garcia, A. Grodsky, R.A. Locarnini, A.V. Mishonov, C.R. Paver, J.R. Reagan, D. Seidov, I.V. Smolyar, K. Weathers, M.M. Zweng,(2018): World Ocean Database 2018. A.V. Mishonov, Technical Ed., NOAA Atlas NESDIS 87. https://www.ncei.noaa.gov/sites/default/files/2020-04/wod_intro_0.pdf

APPENDIX A

Table A1: Additional variables contained in EN4 netCDF profile files.

NetCDF Variable	Description
Global Attributes	<i>Attributes describing the whole file.</i>
History	The history of modifications to this file.
Dimensions	<i>These are the same dimension variables as in an Argo file, although the values given are specifically for EN4.</i>
DATE_TIME	The length of the ASCII date time variable = 14. This is YYYYMMDDHHMiMiSS (where Y=year, M=month, D=day, H=hour, Mi=Minutes and S=seconds).
STRINGV	Maximum length, in characters, of variables that are represented as strings. N can be 2, 4, 8, 16, 64 or 256 in EN4.
N_CALIB	Number of calibrations performed on a variable = 1.
N_PARAM	Maximum number of parameters that could be calculated from a pressure sample = 5. These include depth, potm, psal, bkpt (background potential temperature) and bkps (background practical salinity).
N_HISTORY	The number of history records. History records are masked in EN4, so this is not relevant.
General Information	<i>These variables contain information about the whole profile file. They share the same variable names as those in Argo.</i>
DATA_TYPE	The type of data contained in the file. This was set historically and is always 'ENSEMBLES EN3 v1', no matter which EN version it is.
DATE_CREATION	The date of creation of this file. Using the DATE_TIME format.
DATE_UPDATE	The date of the update of this file. Using DATE_TIME format.
FORMAT_VERSION	The file format version. This is always 2.0.
HANDBOOK_VERSION	The version number of the data handbook. This is always 1.0.

REFERENCE_DATE_TIME	Date of reference for Julian days. This is always 19500101000000 in DATE_TIME format i.e., 1 st January 1950.
General Information for each profile	<i>These are the variables that have an N_PROF dimension. Many share the same variable names as those in Argo, but the contents are different in places.</i>
CYCLE_NUMBER	This is always -2147483647 EN4, but is able to contain the Argo float cycle number.
DATA_CENTRE	The data centre which processed the data. Always MO for Met Office.
DATA_MODE	This is always set to 'D' in EN4. The status of the Argo data (real time, adjusted, or delayed mode) is stored as the fifth character of 'PROJECT_NAME' if the first four characters are 'ARGO'.
DATA_STATE_INDICATOR	This always starts '2C+' following the guidance table in the Argo Data Users' Manual 'Data state indicator recommended use'. The fourth character is 'S' if this is a superob report.
DC_REFERENCE	In EN4 this always begins with the letter A, followed by a nine character code made up of the year, month and day of the observation followed by a '-', followed by the number of this observation in the file – this number can be larger than N_PROF as N_PROF does not count duplicates as they are not kept.
DIRECTION	This is currently always empty in EN4, but is able to contain the Argo profile direction (A for ascending or D for descending).
JULD_LOCATION	A repeat of JULD.
JULD_QC	Quality flag of JULD. Always empty in EN4.
PI_NAME	The name of the Met Office scientist who initially produced the EN4 system.
POSITIONING_SYSTEM	Can give the name of the system in charge of positioning float locations. Always empty in EN4.
PROFILE_DEPH_QC	Global quality flag of the depth profile. Always empty in EN4.
STATION_PARAMETERS	List of parameters contained in this profile.
WMO_INST_TYPE	A broad instrument classification where all profiles are given one of the following identifiers: 401 = XBT/MBT/APB (animal data)/Micro-BT 741 = TESAC (includes CTD/STD/Bottle and others) 831 = Profiling floats 820 = Buoys Because these are very broad classifications it is recommended that INST_REFERENCE is used in preference to WMO_INST_TYPE.
Measurements for each profile	<i>These are the variables that have an N_PROF and an N_LEVELS dimension.</i>
DEPH_CORRECTED_QC	Quality on each DEPH_CORRECTED level. Always empty in EN4.

Calibration information for each profile	<i>Calibration information for each parameter for each profile. If no calibration is available then N_CALIB is set to 1, PARAMETER is filled with the list of parameter names, and all values of the calibration section are set to fill values [as in Argo Users' Manual].</i>
CALIBRATION_DATE	Date of the calibration. Always empty in EN4.
SCIENTIFIC_CALIB_EQUATION	The calibration equation applied to the parameter. For depth (deph) this provides details of how the Hanawa XBT corrections could be applied ($X * \text{depth}$), but this will have already been done if necessary.
SCIENTIFIC_CALIB_COEFFICIENT	The coefficient (X) of the equation applied.
SCIENTIFIC_CALIB_COMMENT	Comment about the calibration. Always empty in EN4.
PARAMETER	Name of the calibrated parameter.
History information for each profile	<i>Variables beginning HISTORY_ are entirely masked in EN4 and are therefore not described here.</i>

Table A2: Additional variables present in EN4 analysis files.

NetCDF Variable	Description
Global Attributes	<i>Attributes describing the whole file.</i>
Conventions	This is always CF-1.0, although, most variables are compliant to CF-1.8 standards.
Title	A description of the file, 'Temperature and salinity analysis'.
DSD_entry_id	This is always UKMO-L4HRfnd-GLOB-OSTIA, but not relevant to analysis file users.
institution	The institution which produced the file. Always 'UK Met Office'.
Contact	The person to contact with queries.
GDS_version_id	Always v1.7, but not relevant to analysis file users.
netCDF_version_id	The netCDF version.
Creation_date	The date this file was created.
File_quality_index	This is always zero.
Grid_resolution	1 degree.
History	Empty.
Product_version	This is always 1.0.
Sothernmost_latitude	This is always -90.5, but in EN4 we do not provide the most southerly latitudes where there is no ocean.
Northernmost_latitude	This is always 89.5.
Westernmost_longitude	This is always 0.5.
Easternmost_longitude	This is always 362.5, but longitudes do only go to 360 in the EN files.
Start_date	Always 1 st Jan 2001.
Start_time	Always 00:00:00.
Stop_date	Always 1 st Jan 2001.
Stop_time	Always 00:00:00.