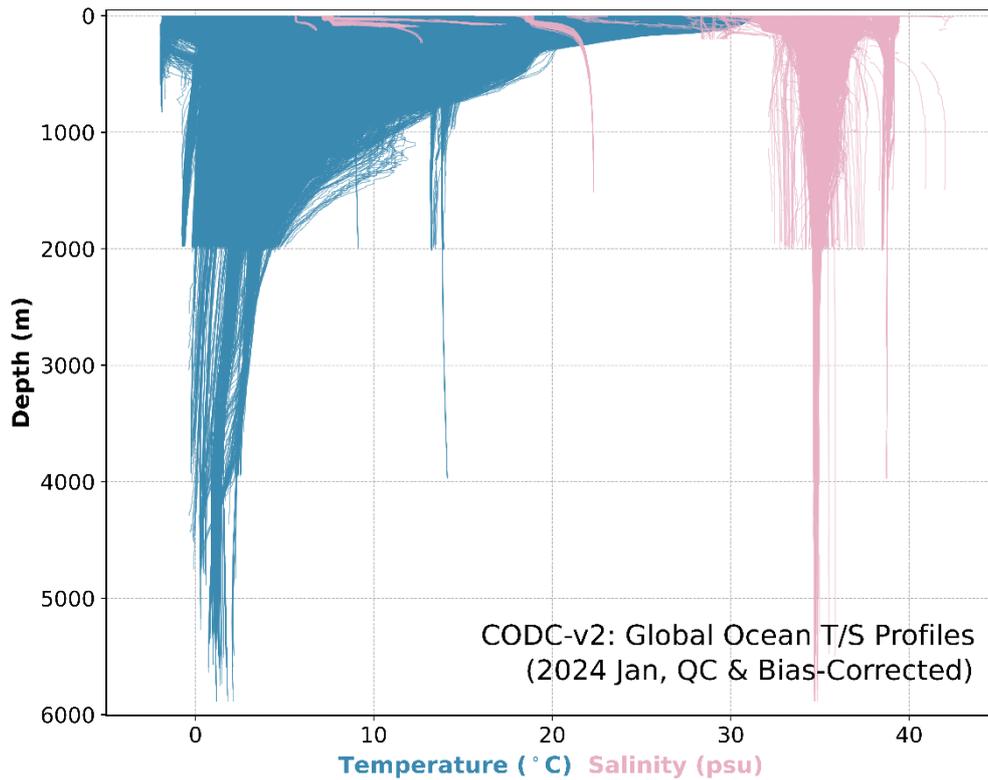




中国科学院  
计算机网络信息中心  
Computer Network Information Center,  
Chinese Academy of Sciences

## CODC-v2 全球海洋温度、盐度廓线数据集



成里京<sup>1</sup> 朱雨静<sup>1</sup> 潘玉莹<sup>1</sup> 谭哲韬<sup>1</sup> 宋欣焱<sup>1</sup>

原惠峰<sup>2</sup> 张斌<sup>3</sup> 朱江<sup>1</sup> 王凡<sup>3</sup>

<sup>1</sup> 中国科学院大气物理研究所

<sup>2</sup> 中国科学院计算机网络信息中心

<sup>3</sup> 中国科学院海洋研究所

# 中国科学院海洋科学数据中心

## 全球海洋科学数据集 (CODC)

### 温盐廓线数据说明(MAT 格式/NetCDF-4 格式)

第二版本的 CODC 全球海洋温度、盐度廓线科学数据集 (CAS-Oceanographic Data Center, Global Ocean Science Database: CODC-v2) 由中国科学院海洋科学数据中心共建单位大气物理研究所牵头, 联合中国科学院海洋研究所、中国科学院计算机网络信息中心构建, 主要提供经过质量控制、偏差订正后的海洋温度、盐度廓线数据。该数据集涵盖自 1940 年以来的观测数据, 数据包括 XBT、CTD、Argo、滑翔机 (Glider)、浮标等 14 种观测仪器, 数据量为: 温度 18472149 条(图 1)、盐度 11273326 条 (图 2) (数据集时间段: 1940 年 1 月-2024 年 11 月)。

## 一、数据文件说明

**数据源:** CODC-v2 的数据源以 WOD (World Ocean Database) 为主, GTSP (Global Temperature and Salinity Profile Programme) 为辅。由于 WOD 数据每季度更新, GTSP (最佳版本 best copy) 数据每月更新, 当 WOD 数据源未更新时, CODC-v2 会使用 GTSP 数据源作为补充, 不同数据源生成的 CODC-v2 数据文件内容一致, 但文件名不同: WOD 数据源文件以“WOD\_”开头, GTSP 数据源文件以“GTSP\_”开头。

**数据文件格式:** CODC-v2 将所有观测数据按月汇总, 对原始

盐度数据进行质量控制、对原始温度数据进行质量控制和多种仪器偏差订正,生成 1940 年 1 月至今的数据。数据提供两种格式: MAT 和 NetCDF-4, 同一月份的两种格式文件内容一致。**建议用户优先使用经过质量控制和偏差订正的温盐数据 (即 Temperature、Depth\_Temperature 和 Salinity、Depth\_Salinity 变量)。**

**数据更新频率:** 滞后 7-30 天更新上个月的数据。

**数据更新方式:** 由于历史数据会有滚动更新,例如 Argo 数据会不断有延时质控数据替换实时数据,因此 CODC-v2 也会对过去几个月的数据进行滚动更新,并在每年 1 月对上一年全部数据进行整体更新。

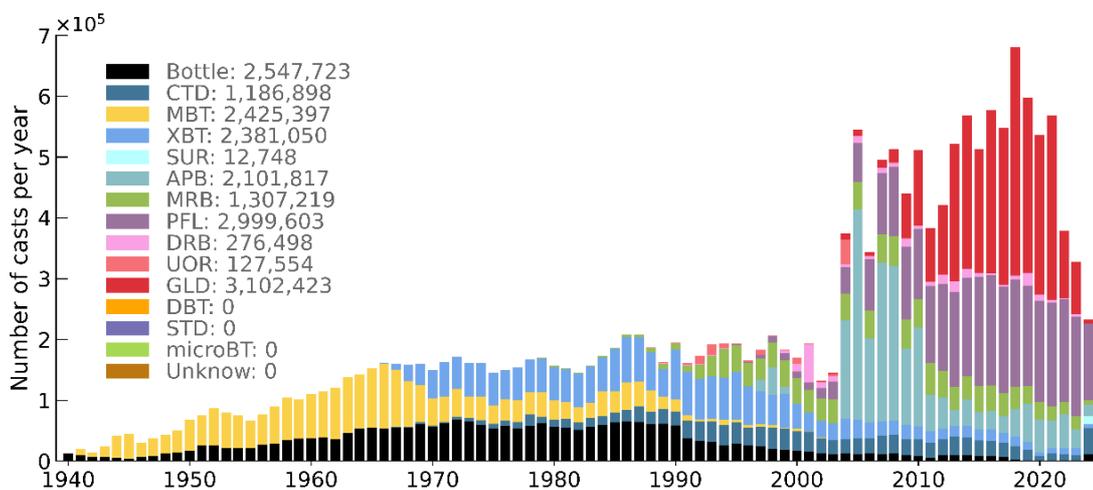


图 1: 历年温度廓线数量统计。

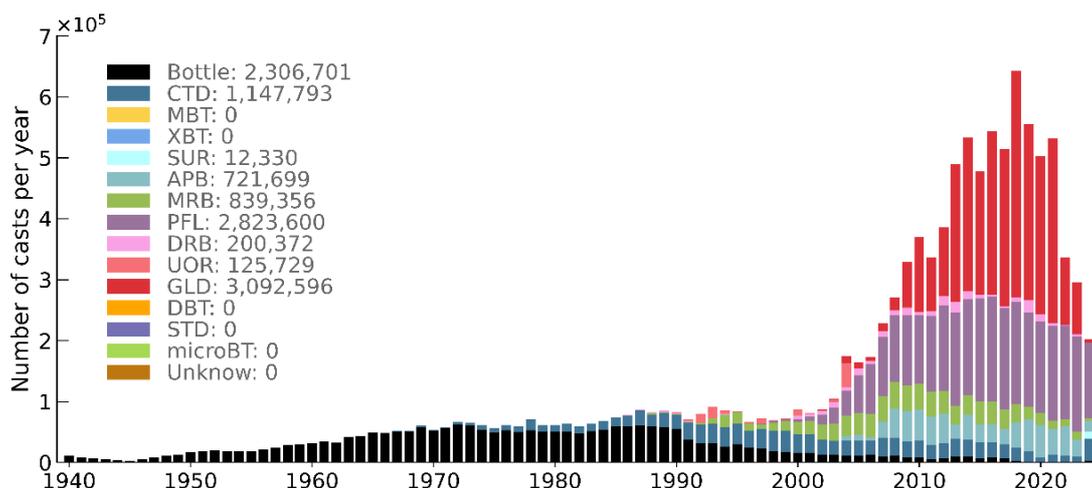


图 2：历年盐度廓线数量统计。

### (一) 变量一览表

变量名称	维度	数据类型
Tprofile_num_all	[1, 1]	int32
Profile_info_record_name	[18, 1]	string
Profile_info_record_all	[18, Tprofile_num_all]	float64
Profile_info_str_name	[18, 1]	string
Profile_info_str_all	[18, Tprofile_num_all]	string
Depth_origin	[3000 <sup>①</sup> , Tprofile_num_all]	float64
Temp_origin	[3000, Tprofile_num_all]	float64
Salinity_origin	[3000, Tprofile_num_all]	float64
Temperature	[3000, Tprofile_num_all]	float64
Depth_Temperature	[3000, Tprofile_num_all]	float64
Salinity	[3000, Tprofile_num_all]	float64
Depth_Salinity	[3000, Tprofile_num_all]	float64
Depth_XBT	[3000, Tprofile_num_all]	float64

<sup>①</sup> 每列表示一条廓线的观测点数据，每列的 3000 行数据对应该廓线从海表到海底的观测点。由于自 1940 年以来，仅有不到 0.26% 的廓线所观测的点数超过 3000。为了保留尽可能多的观测点且满足存储限制，CODCv2 变量的最多深度层数设定为 3000。

Temp_XBT	[3000, Tprofile_num_all]	float64
Depth_APB	[3000, Tprofile_num_all]	float64
Temp_APB	[3000, Tprofile_num_all]	float64
Depth_MBT	[3000, Tprofile_num_all]	float64
Temp_MBT	[3000, Tprofile_num_all]	float64
Depth_BOT	[3000, Tprofile_num_all]	float64
Temp_BOT	[3000, Tprofile_num_all]	float64
Temperature_CODCflag	[3000, Tprofile_num_all]	int16
Temperature_CODCflag_checks	[3000, Tprofile_num_all, 15]	int16
Salinity_CODCflag	[3000, Tprofile_num_all]	int16
Salinity_CODCflag_checks	[3000, Tprofile_num_all, 13]	int16
Depth_WODflag <sup>②</sup>	[3000, Tprofile_num_all]	int16
Temperature_WODflag	[3000, Tprofile_num_all]	int16
Salinity_WODflag	[3000, Tprofile_num_all]	int16

\* 上述的变量维度是指，在 MATLAB 中读取 MAT 格式和在 Python 中读取 NetCDF-4 格式得到的变量维度，如果使用其他软件读取变量，变量维度可能会发生翻转。

## (二) 变量详细说明

1. **Tprofile\_num\_all**: 该月份的观测廓线总数量 (\*Tprofile 代表 Total profiles)。
2. **Profile\_info\_record\_all**: 存储了每条廓线的 18 个、可用浮点数表达的元数据信息，每列代表一条廓线的 18 个信息，每行具体信息见下述变量 Profile\_info\_record\_name。
3. **Profile\_info\_record\_name**: 解释 Profile\_info\_record\_all 中每行信息的含义
  - 1) WOD\_unique\_id: WOD 唯一标识符
  - 2) Year: 年
  - 3) Month: 月

<sup>②</sup> 如果数据源为 GTSP, 数据文件中深度质控符对应的变量为 Depth\_GTSPflag。同理，温度和盐度质控结果分别存储在 Temperature\_GTSPflag 和 Salinity\_GTSPflag 中。

- 4) Day: 日
- 5) Latitude: 纬度 (单位: 北纬为正, 南纬为负)
- 6) Longitude: 经度 (单位: 东经为正, 西经为负)
- 7) Instrument type: 仪器类型代码, 数值的具体含义为: 1=OSD; 2=CTD; 3=MBT; 4=XBT; 5=SUR; 6=APB; 7=MRB; 8=PFL; 9=DRB; 10=UOR; 11=GLD; 12=DBT; 13=STD; 14=microBT; -999=Unknown<sup>③</sup>。
- 8) GMT\_time: 格林尼治标准时间, 一天中从午夜 (0:00) 开始的小时数。
- 9) Launch height: (XBT) 仪器从船舶上的投放高度。
- 10) original station number: 数据来源的站点编号或标识符。
- 11) access number: 接收编号, 是一个唯一的编号, 用于标识一组海洋数据, 通常由 NCEI (National Centers for Environmental Information) 分配, 并用于长期存档。每当新的海洋数据被提交给 NCEI 存档时, 这些数据会被赋予一个唯一的 Accession Number, 通过这个编号, 用户可以访问原始数据以及与这些数据相关的信息。
- 12) WMO\_id: 由世界气象组织 (WMO) 分配给浮标或剖面浮筒的代码。
- 13) Time(days since 1770-01-01 00:00:00): 时间信息 (自 1770-1-1 00:00:00 起始的时间)。
- 14) Temperature\_WODprofileflag<sup>④</sup>: 该廓线在 WOD 数据集中的温度质控符 (廓线整体的质控符)。
- 15) Salinity\_WODprofileflag<sup>⑤</sup>: 该廓线在 WOD 数据集中的盐度质控符 (廓线整体

---

<sup>③</sup> 缩写含义:

OSD: Ocean Station Data, Low-resolution CTD/XCTD, Plankton data  
 CTD: High-resolution Conductivity-Temperature-Depth / XCTD data  
 MBT: Mechanical / Digital / Micro Bathythermograph data  
 XBT: Expendable Bathythermograph data  
 SUR: Surface-only data  
 APB: Autonomous Pinniped data  
 MRB: Moored buoy data  
 PFL: Profiling float data  
 DRB: Drifting buoy data  
 UOR: Undulating Oceanographic Recorder data  
 GLD: Glider data  
 DBT: Digital Bathythermograph  
 micro BT: Micro Bathythermograph

<sup>④</sup> 如果数据源为 GTSP, 数据文件该信息对应为 Temperature\_GTSPprofileflag。

<sup>⑤</sup> 如果数据源为 GTSP, 数据文件该信息对应为 Salinity\_GTSPprofileflag。

的质控符)。

- 16) **Temperature Adjustment:** 对原始温度测量值进行的调整。由 Argo 项目数据管理团队对 Argo 浮标进行的数据调整值 (例如: 修正、校准)。该调整值为实数 (即小数), 表示原始 (实时, **real-time**) 和调整后的 (延迟模式, **delayed-mode**) 温度、盐度、压力廓线数据之间的平均差异, 适用于 500 米深度以下的所有值。如果一个剖面有调整值 (即使该值为 0.0), 也表明该剖面已通过 Argo 项目的额外质量控制, 并被视为调整后的实时数据或延迟模式数据。
- 17) **Salinity Adjustment:** 对原始盐度测量值进行的调整, 具体含义见 **Temperature Adjustment**。
- 18) **Pressure Adjustment:** 对原始压力测量值进行的调整, 具体含义见 **Temperature Adjustment**。

**4. Profile\_info\_str\_all:** 存储了每条廓线的 18 个、可用字符表达的元数据信息, 每列代表一条廓线的 18 个信息, 每行具体信息见下述变量 **Profile\_info\_str\_name**。

**5. Profile\_info\_str\_name:** 解释 **Profile\_info\_str\_all** 中每行信息的含义

- 1) **CAS\_id:** IAP 用于追踪原始数据的唯一性 id。
- 2) **Temperature instrument:** 温度测量设备。
- 3) **Salinity instrument:** 盐度测量设备。
- 4) **Salinity/Temperature recorder:** 数据记录系统。
- 5) **Need\_z\_fix\_Instrument (dpc):** WOD 建议的深度订正标记。
- 6) **WOD\_cruise\_identifier:** 国家代码和 WOD 的船舶识别码组成的标识。
- 7) **Country:** 国家信息。
- 8) **database\_orig:** 指示数据的来源数据库。
- 9) **Platform:** 数据采集时使用的平台的具体名称。
- 10) **real\_time\_data:** 标识是否为通过 GTS 快速传输系统在 24 小时内收到的实时数据。
- 11) **originators\_cruise\_identifier:** 由原始数据提供者定义的航次编号。
- 12) **Project:** 项目名称, 标识数据所属的研究或调查项目。
- 13) **Ocean\_Vehicle:** 海洋观测载具, 指的是用于采集海洋环境数据的各种仪器或设

备。

14) **Institute**: 指采集数据的机构或组织。

15) **XBT\_depth\_eq**: XBT 下降方程。

16) **Temperature data mode**: 温度数据所处的处理和发布进度的阶段(包括实时数据 **real-time**、延时数据 **delayed**、实时调整数据 **adjusted real-time**; 主要是 Argo 数据)。

17) **Salinity data mode**: 盐度数据所处的处理和发布进度的阶段(包括实时数据 **real-time**、延时数据 **delayed**、实时调整数据 **adjusted real-time**; 主要是 Argo 数据)。

18) **Pressure data mode**: 压力数据所处的处理和发布进度的阶段(包括实时数据 **real-time**、延时数据 **delayed**、实时调整数据 **adjusted real-time**; 主要是 Argo 数据)。

6. **Depth\_origin**: 原始深度数据(未经质量控制和仪器偏差订正)。

7. **Temp\_origin**: 原始温度数据(未经质量控制和仪器偏差订正)。

8. **Salinity\_origin**: 原始盐度数据(未经质量控制和仪器偏差订正)。

9. **Temperature<sup>⑥</sup>**: 经温度质量控制和仪器偏差订正(XBT、MBT、Bottle、APB)后的温度廓线数据。

10. **Depth\_Temperature**: 与 **Temperature** 对应的深度数据。

11. **Salintiy<sup>⑦</sup>**: 经质量控制后的盐度廓线数据。

12. **Depth\_Salinity**: 与 **Salinity** 对应的深度数据。

13. **Depth\_XBT**、**Temp\_XBT**: 使用 CH14 (Cheng et al., 2014) 方法订正后的 XBT 深度和温度数据, 对于非 XBT 仪器观测得到的廓线数据, 其对应的列填充为 NaN (可通过 **Instrument type** 提取所有 XBT 数据, 详见(三)基本数据使用示例)。

14. **Depth\_APB**、**Temp\_APB**: 使用自主研发的 APB 方法 (Gouretski, Roquet, and Cheng, 2024) 订正后的 APB 深度的温度数据, 对于非 APB 仪器观测得到的廓线数据, 其对应的列填充为 NaN (可通过 **Instrument type** 提取所有 APB 数

<sup>⑥</sup> 建议用户使用经过温度质量控制和仪器偏差订正的温度数据及对应的深度数据。

<sup>⑦</sup> 建议用户使用经过盐度质量控制的盐度数据和对应的深度数据。

据，详见（三）基本数据使用示例）。

**15. Depth\_MBT、Temp\_MBT:** 使用 GC20 (Gourestki and Cheng, 2021) 方法订正后的 MBT 深度和温度数据，对于非 MBT 仪器观测得到的廓线数据，其对应的列填充为 NaN (可通过 Instrument type 提取所有 MBT 数据，详见（三）基本数据使用示例)。

**16. Depth\_BOT、Temp\_BOT:** 使用 GC22 (Gourestki and Cheng, 2022) 方法订正后的 Bottle 深度和温度数据，对于非 Bottle 仪器观测得到的廓线数据，其对应的列填充为 NaN (可通过 Instrument type 提取所有 Bottle 数据，详见（三）基本数据使用示例)。

**17. Temperature\_CODCflag:** CODC-QC 的温度总质控符，其中 0 为通过了 15 个温度质量检查模块的、可接受的数据，1 为至少有 1 个质量检查模块未通过的、拒绝的数据。每一个观测点都对应有一个总质控符，即总质控符与 Temp\_origin 内的观测点一一对应。

**18. Temperature\_CODCflag\_checks:** CODC-QC 质控系统中 15 个 (对应该变量的第三维) 温度质量检查模块的检查结果，0 为通过该质量检查模块，1 为未通过该质量检查模块，缺失值由 -999 表示。

**19. Salinity\_CODCflag:** CODC-QC 的盐度总质控符，其中 0 为通过了 13 个盐度质量检查模块的、可接受的数据，1 为至少有 1 个质量检查模块未通过的、拒绝的数据。每一个观测点都对应有一个总质控符，即总质控符与 Salinity\_origin 内的观测点一一对应。

**20. Salinity\_CODCflag\_checks:** CODC-QC 质控系统中 13 个 (即该变量的第三维) 盐度质量检查模块的检查结果，0 为通过质量检查模块，1 为未通过该质量检查模块，缺失值由 -999 表示。

**21. Depth\_WODflag<sup>®</sup>:** NOAA/NCEI 提供的深度质控符，0 表示质量良好数据，

---

<sup>®</sup> 如果数据源为 GTSP，数据文件中，数据源提供的深度质控符对应的变量为 Depth\_GTSPflag，该变量由 GTSP 提供，使用不同的标志值 (0-9) 表示数据的质量状态，如“良好”、“可能有问题”或“缺失”。详细定义可参考 <https://www.nodc.noaa.gov/GTSP/document/qcmans/qcflags.htm>。同理，温度和盐度的质控变量分别为 Temperature\_GTSPflag 和 Salinity\_GTSPflag，其含义与 Depth\_GTSPflag 相同。

非 0 表示质量可疑数据，每一个观测点都对应 1 个质控符。详细定义可参考 <https://www.ncei.noaa.gov/access/world-ocean-database/CODES/definition-quality-flags.html>

**22. Temperature\_WODflag:** NOAA/NCEI 提供的温度质控符，0 表示质量良好数据，非 0 表示质量可疑数据，每一个观测点都对应 1 个质控符。详细定义可参考 <https://www.ncei.noaa.gov/access/world-ocean-database/CODES/definition-quality-flags.html>

**23. Salinity\_WODflag:** NOAA/NCEI 提供的盐度质控符，0 表示质量良好数据，非 0 表示质量可疑数据，每一个观测点都对应 1 个质控符。详细定义可参考 <https://www.ncei.noaa.gov/access/world-ocean-database/CODES/definition-quality-flags.html>

### (三) 基本数据使用示例

CODC-v2 提供的 MAT 格式基于 HDF5 格式(MATLAB 7.3 及以上版本)，无法使用传统的 load 命令读取。读取和处理 MAT 文件需要 MATLAB R2021a 或更高版本（支持 HDF5 格式功能）。在此提供了 MATLAB 中读取 MAT 文件和 Python 中读取 NetCDF-4 文件的示例代码，并展示了数据的基本处理方法，便于用户了解如何使用相关数据。由于 MAT 文件体积较大，可能占用较多内存或加载时间，**因此建议用户优先使用 NetCDF-4 文件。**

#### 1. MATLAB 读取 MAT 文件

```
FileName='WOD_CAS_T_S_2023_6.mat'  
% 查看 MAT 文件中的内容（结构和变量）  
h5disp(FileName)  
% 用于获取 HDF5 文件的元数据，包括文件中的数据、组和属性  
h5info(FileName)  
% 读取数据
```

```

Temp_origin = h5read(FileName, '/Temp_origin');
Temperature_CODCflag = h5read(FileName, '/Temperature_CODCflag');
Temperature_afterQC = Temp_origin;
% 利用质控符保留通过质控的数据
Temperature_afterQC(Temperature_CODCflag ~= 0) = NaN;
Depth_origin = h5read(FileName, '/Depth_origin');
Depth_afterQC = Depth_origin;
Depth_afterQC(Temperature_CODCflag~=0) = NaN;
% 获取仪器信息
Profile_info_record_all = h5read(FileName, '/Profile_info_record_all');
instrument_type = Profile_info_record_all(7,:);
% 提取所有 PFL 数据
Temp_PFL = Temp_origin;
Temp_PFL = Temp_PFL(:, instrument_type == 8);
% 提取所有经过 XBT 偏差订正后的 XBT 数据
Temp_XBTcor = h5read(FileName, '/Temp_XBTcor');
Depth_XBTcor = h5read(FileName, '/Depth_XBTcor');
Temp_XBTcor = Temp_XBTcor(:, instrument_type == 4);
Depth_XBTcor = Depth_XBTcor(:, instrument_type == 4);
% 提取 XBT 数据对应的年月日和经纬度信息
Year= Profile_info_record_all(2, instrument_type == 4);
Month= Profile_info_record_all(3, instrument_type == 4);
Day= Profile_info_record_all(4, instrument_type == 4);
Latitude= Profile_info_record_all(5, instrument_type == 4);
Longitude= Profile_info_record_all(6, instrument_type == 4);

```

## 2. Python 读取 NetCDF-4 文件

```

import netCDF4 as nc
import copy
FileName='WOD_CAS_T_S_2023_6.nc'
dataset=nc.Dataset(FileName)
# 查看 NetCDF-4 文件中的内容
print(dataset)
# 读取数据
Temp_origin=dataset.variables['Temp_origin'][:].data
Temperature_CODCflag=dataset.variables['Temperature_CODCflag'][:].data
Depth_origin=dataset.variables['Depth_origin'][:].data
# 利用质控符保留通过质控的数据

```

```

Temperature_afterQC = copy.deepcopy(Temp_origin)
Temperature_afterQC[Temperature_CODCflag != 0] = np.nan
Depth_afterQC = copy.deepcopy(Depth_origin)
Depth_afterQC[Temperature_CODCflag != 0] = np.nan
# 获取仪器信息
Profile_info_record_all=dataset.variables['Profile_info_record_all'][:].data
instrument_type = Profile_info_record_all[6, :]
# 提取所有 PFL 数据
Temp_PFL = Temp_origin[:, instrument_type == 8]
# 提取所有经过 XBT 偏差订正后的 XBT 数据
Temp_XBTcor = dataset.variables['Temp_XBTcor'][:].data
Depth_XBTcor = dataset.variables['Depth_XBTcor'][:].data
Temp_XBTcor = Temp_XBTcor[:, instrument_type == 4]
Depth_XBTcor = Depth_XBTcor[:, instrument_type == 4]
# 提取 XBT 数据对应的年月日和经纬度信息
Year= Profile_info_record_all[2, instrument_type == 4]
Month= Profile_info_record_all[3, instrument_type == 4]
Day= Profile_info_record_all[4, instrument_type == 4]
Latitude= Profile_info_record_all[5, instrument_type == 4]
Longitude= Profile_info_record_all[6, instrument_type == 4]

```

#### (四) 更新说明及注意事项

1. 相对于 CODC-v1 版本，新版本的更新包括：（1）温度、盐度质量控制方案略有更新，解决了 v1 版本中的一些问题；（2）更新了 Bottle 和 APB 数据的偏差订正；XBT 数据偏差订正方案更新到了 2024 版本；（3）v1 中数据源为 WOD 和自有数据，v2 新增加了 GTSP 的数据，加快了 v2 的更新频率。
2. 目前温度、盐度数据的深度是区分开的，主要是因为温度-深度廓线在进行偏差订正时修改了深度数据，但该订正还未在盐度数据中进行充分验证，下一个版本的 CODC 数据会对温盐数据的深度进行统一。

3. 当原始廓线的观测值超过 3000 个时，首先根据数据源提供的质控符去除可疑数据。如果剩余的观测值仍超过 3000 个，将在确保保留最浅和最深深度对应的观测值的前提下，均匀地选取 3000 个观测点，保留这些温度或盐度观测值及其对应的深度。

## 二、参考文献

**CODCv2 数据引用: Yujing Zhu, Yuying Pan, Zhetao Tan, Xinyi Song, Huifeng Yuan, Bin Zhang, Jiang Zhu, Fan Wang, Lijing Cheng, CODC-v2 global ocean in-situ profile observational dataset. Oceanographic Data Center, Chinese Academy of Sciences (<https://msdc.qdio.ac.cn>),2024. DOI: 10.12157/IOCAS.20241217.001.**

1. Cheng, L., Zhu, J., Cowley, R., Boyer, T. & Wijffels, S. Time, Probe Type, and Temperature Variable Bias Corrections to Historical Expendable Bathythermograph Observations. *J. Atmospheric Ocean. Technol.* **31**, 1793–1825 (2014).
2. Gouretski, V. & Cheng, L. Correction for Systematic Errors in the Global Dataset of Temperature Profiles from Mechanical Bathythermographs. *J. Atmospheric Ocean. Technol.* **37**, 841–855 (2020).
3. Gouretski, V., Cheng, L. & Boyer, T. On the Consistency of the Bottle and CTD Profile Data. *J. Atmospheric Ocean. Technol.* **39**, 1869–1887 (2022).
4. Gouretski, V., Roquet, F. & Cheng, L. Measurement Biases in Ocean Temperature Profiles from Marine Mammal Dataloggers. *J. Atmospheric Ocean. Technol.* **41**, 629–645 (2024).
5. Tan, Z. *et al.* A new automatic quality control system for ocean profile observations and impact on ocean warming estimate. *Deep Sea Res. Part Oceanogr. Res. Pap.* **194**, 103961 (2023).
6. Zhang, B. *et al.* CODC-v1: a quality-controlled and bias-corrected ocean temperature profile database from 1940–2023. *Sci. Data* **11**, 666 (2024).

### 三、 联系人

1. IAP 数据专用邮箱, iapocean\_data@mail.iap.ac.cn
2. 成里京, chenglij@mail.iap.ac.cn
3. 潘玉莹, panyuying@mail.iap.ac.cn
4. 朱雨静, zhuyujing22@mails.ucas.ac.cn

### 四、 数据引用说明

**Yujing Zhu, Yuying Pan, Zhetao Tan, Xinyi Song, Huifeng Yuan, Bin Zhang, Jiang Zhu, Fan Wang, Lijing Cheng, CODC-v2 global ocean in-situ profile observational dataset. Oceanographic Data Center, Chinese Academy of Sciences (<https://msdc.qdio.ac.cn>),2024. DOI: 10.12157/IOCAS.20241217.001.**