Manual of Global Ocean Argo gridded data set (BOA_Argo)
(Version 2020)

Shaolei Lu[1]，Zenghong Liu[1]，Hong Li[2]，Zhaoqin Li[1]，Xiaofen Wu[1]，Chaohui Sun[1]，Jianping Xu[1]

1. State Key Laboratory of Satellite Ocean Environment Dynamics (SOED),Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China
2. Ministry of Education Key Laboratory for Earth System Modeling, and Department of Earth System Science, Tsinghua University, Beijing 100084, China

Hangzhou, China

May 2020

This document should be cited as:

Shaolei Lu，Zenghong Liu，Hong Li，Zhaoqin Li，Xiaofen Wu，Chaohui Sun，Jianping Xu.(2020). Manual of Global Ocean Argo gridded data set (BOA_Argo) (Version 2019), 14 pp.

For details about the development of BOA_Argo data set, please refer to the paper " Li, H., F. Xu, W. Zhou, D. Wang, J. S. Wright, Z. Liu, and Y. Lin (2017), Development of a global gridded Argo data set with Barnes successive corrections, *J. Geophys. Res.Oceans*, 122, doi: 10.1002/2016JC012285.6".

Contents

1. History

Table 1 provides a detailed description about the history of BOA_Argo gridded data set.

2. Introduction

The Array for Real-time Geostrophic Oceanography (Argo) program was originally proposed by scientists from the United States and Japan at the end of the 20th century. By the end of 2007, the global Argo network consisted of 3000 Lagrangian free-drifting profiling floats had been accomplished at a spacing of approximately 300 km (3°×3°). The target of the network is to provide continuous, high-resolution temperature and salinity (T/S) observations within the upper 2000 m of the global ocean. Argo data are made available in near real-time, within a few hours after data collection. Over two million Argo temperature-salinity profiles had been obtained by November 2018, from then on, more than 15000 additional profiles per year have been added. Both the data coverage and volume exceed previous traditional observations, however, like those previous observations, temperature and salinity profiles from Argo floats have an irregular spatial and temporal distribution, thus confines its applications.

Many Argo member states have developed their own individual monthly gridded Argo products (e.g. Jamstec-Argo, *Hosoda et al.*, 2008; Roemmich-Argo, *Roemmich and Gilson*, 2009; EN4-Argo, *Good et al.*, 2013; IPRC-Argo, http://apdrc.soest.hawaii.edu/projects/Argo/data/Documentation/gridded-var.pdf). These Argo gridded data sets are mostly based on optimum interpolation (OI) or more sophisticated variational analysis methods (*Troupin et al.*, 2010). As pointed out by *Locarnini et al*. (2013), OI is based on the second-order statistics, which tends to induce large errors when data is relatively sparse.

The advantage of Barnes corrections is that it is able to capture signals at

Table 1. History of the BOA_Argo gridded data set

Version	Domain	Data source	First guess	Objective analysis method	Spatial resolution	Vertical resolution	Temporal coverage , resolution	Variable	Online time
BOA_Argo 2012	Global ocean (180 °W-180 °E, 59.5 °S-59.5 °N）	Argo T/S	Cressman(1955), Argo 2004-2011 T/S climatology data	Barnes (1973), new respond function	1 degree	48 levels, 5-1950m	2004-2011, monthly	T/S	08/2012 (*Li et al*.,2012)
BOA_Argo 2013	Global ocean (180 °W-180 °E, 59.5 °S-59.5 °N）	Argo T/S	Argo 2004-2012 T/S climatology data	The same as BOA_Argo2012	1 degree	48 levels, 5-1950m	2004-2012, monthly	T/S	04/2013 (*Li et al*.,2013)
BOA_Argo 2014	Global ocean (180 °W-180 °E, 59.5 °S-59.5 °N）	Argo T/S	Argo 2004-2013 T/S climatology data	The same as BOA_Argo2013	1 degree	48 levels, 5-1950m	2004-2013, monthly	T/S	04/2014 (*Li et al*.,2014)
BOA_Argo 2015	Global ocean (180 °W-180 °E, 59.5 °S-59.5 °N）	Argo T/S	Argo 2004-2014 T/S/ ILD/MLD/CMLD climatology data	The same as BOA_Argo2014	1 degree	49 levels, 0-1950m	2004-2014, monthly	T/S/ILD/ MLD/CMLD[1]	06/2015 (*Li et al*.,2015)
BOA_Argo 2016	Global ocean (180 °W-180 °E, 79.5 °S-79.5 °N）	Argo T/S	Argo 2004-2015 T/S/ ILD/MLD/CMLD climatology data	The same as BOA_Argo2015	1 degree	58 levels, 0-1975dbar	2004-2015, monthly	T/S/ILD/ MLD/CMLD	06/2016 (*Li et al*.,2016)

---

[1] **T**: temperature, **S**: salinity, **ILD**: Isothermal Layer Depth, **MLD**: Mixed Layer Depth, **CMLD**: Composed Mixed Layer Depth.

BOA_Argo 2017	Global ocean (180 °W-180 °E, 79.5 °S-79.5 °N）	Argo T/S	Argo 2004-2016 T/S/ ILD/MLD/CMLD climatology data	The same as BOA_Argo2016	1 degree	58 levels, 0-1975dbar	2004-2016, monthly	T/S/ILD/ MLD/CMLD	05/2017 (*Li et al.*,2017)
BOA_Argo 2018	Global ocean (180 °W-180 °E, 79.5 °S-79.5 °N）	Argo T/S	Argo 2004-2017 T/S/ ILD/MLD/CMLD climatology data	The same as BOA_Argo2017	1 degree	58 levels, 0-1975dbar	2004-2017, monthly	T/S/ILD/ MLD/CMLD	06/2018 (*Lu et al.,2018*)
BOA_Argo 2019	Global ocean (180 °W-180 °E, 79.5 °S-79.5 °N）	Argo T/S	Argo 2004-2018 T/S/ ILD/MLD/CMLD climatology data	The same as BOA_Argo2018	1 degree	58 levels, 0-1975dbar	2004-2018, monthly	T/S/ILD/ MLD/CMLD	06/2019 (*Lu et al.,2019*)
BOA_Argo 2020	Global ocean (180 °W-180 °E, 79.5 °S-79.5 °N）	Argo T/S	Argo 2004-2019 T/S/ ILD/MLD/CMLD climatology data	The same as BOA_Argo2019	1 degree	58 levels, 0-1975dbar	2004-2019, monthly	T/S/ILD/ MLD/CMLD	06/2020 (*This manual*)

various scales based on pre-defined response functions. However, there has been no monthly Argo gridded data set generated by the Barnes method so far, because the classic Barnes method also requires uniformly distributed observations in spatial.

Levitus (1982) successfully developed a global climatology of ocean temperature and salinity using simple Barnes successive corrections. Theoretically, this enables the production of objective analysis via examinations of the response function and calculation of the key parameters. However, the Argo monthly data are not evenly distributed and relatively sparse compared with WOA observational data. This poses a significant obstacle to apply the default Barnes correction method directly to the irregularly distributed monthly Argo data.

We refine the classic Barnes method by adopting the most suitable parameters and response functions. A series of error analysis are conducted to quantify the most appropriate parameters in the Barnes method, including the iteration number, influence radius, convergence factor, and filtering parameter. New response functions are then derived based on these parameters. Compared to the three iterations required in the classic Barnes method, the new approach only requires two iterations. Due to the flexible response functions used, the refined method is able to retain more mesoscale signals than the classic method.

Then a new Argo gridded data set (Barnes objective analysis, BOA_Argo) is produced following the below procedures. First, a quality control filter is used for screening all available Argo profiles and, the geometry model method proposed by *Chu et al.* (2011) is employed to calculate the Mixed Layer Depth (MLD) for each Argo profile. The corresponding sea surface (0dbar and 5dbar) temperature and salinity are then estimated using the linear fitting method in the mixed layer. The initial background state is constructed using Cressman successive correction (*Cressman,* 1957). The Barnes successive correction is then applied to produce monthly gridded analyses. We use this approach to generate the monthly mean 3-dimensional temperature, salinity, and derived products in the global ocean.

3. Data preparing

The global Argo temperature and salinity profiles are provided by the China Argo Real-time Data Center (http://www.argo.org.cn). All of the data have been quality controlled (either a real-time quality control (QC) or a delayed-mode QC) at the national Argo Data Assembly Centers (DACs). Users should refer to the Argo data management documentation (Wong et al., 2015, Argo quality control manual for CTD and trajectory data Version 3.0, http://dx.doi.org/10.13155/33951) for details.

Despite the above QC criteria, the quality of some data is still insufficient for our application. We further improve the quality and suitability of the profiles ingested into BOA_Argo by applying some additional filters that listed in Figure 1.

4. Scheme of production of the BOA_Argo gridded data set

The generation of BOA_Argo data follows the four steps (Figure 1).

(1) Data screening

The Argo data quality control procedure is as shown by Figure 1, which yields data on a uniform 58 levels in vertical. Figure 2 shows the time series of monthly number of the profiles before and after the application of our quality control criteria between January 2004 and December 2019.

(2) Construction of the first guess field

A monthly climatological initial conditions are obtained via Cressman analysis. *Cressman* (1959) introduced a successive correction scheme, which iteratively corrects gridded background values (first guesses) by applying a linear combination of corrections between predicted and observed values. Corrections are applied at each grid point using the equation

$$f_i^{n+1} = f_i^n + \frac{\sum_{b=1}^{K_i^n} w_{ib}^n (f_b^o - f_b^n)}{\sum_{b=1}^{K_i^n} w_{ib}^n}, \qquad (1)$$

where f_i^n is the analysis field after the n^{th} iteration at grid point i, f_b^o is the b^{th} observation within the scan radius R_n (called the influence region) for the n^{th} iteration, f_b^n is the estimate after the n^{th} iteration interpolated to the location of the b^{th} observation, K_i^n is the number of the observations collected within R_n, and w_{ib}^n is a weighting function. The weighting function w_{ib}^n is defined as:

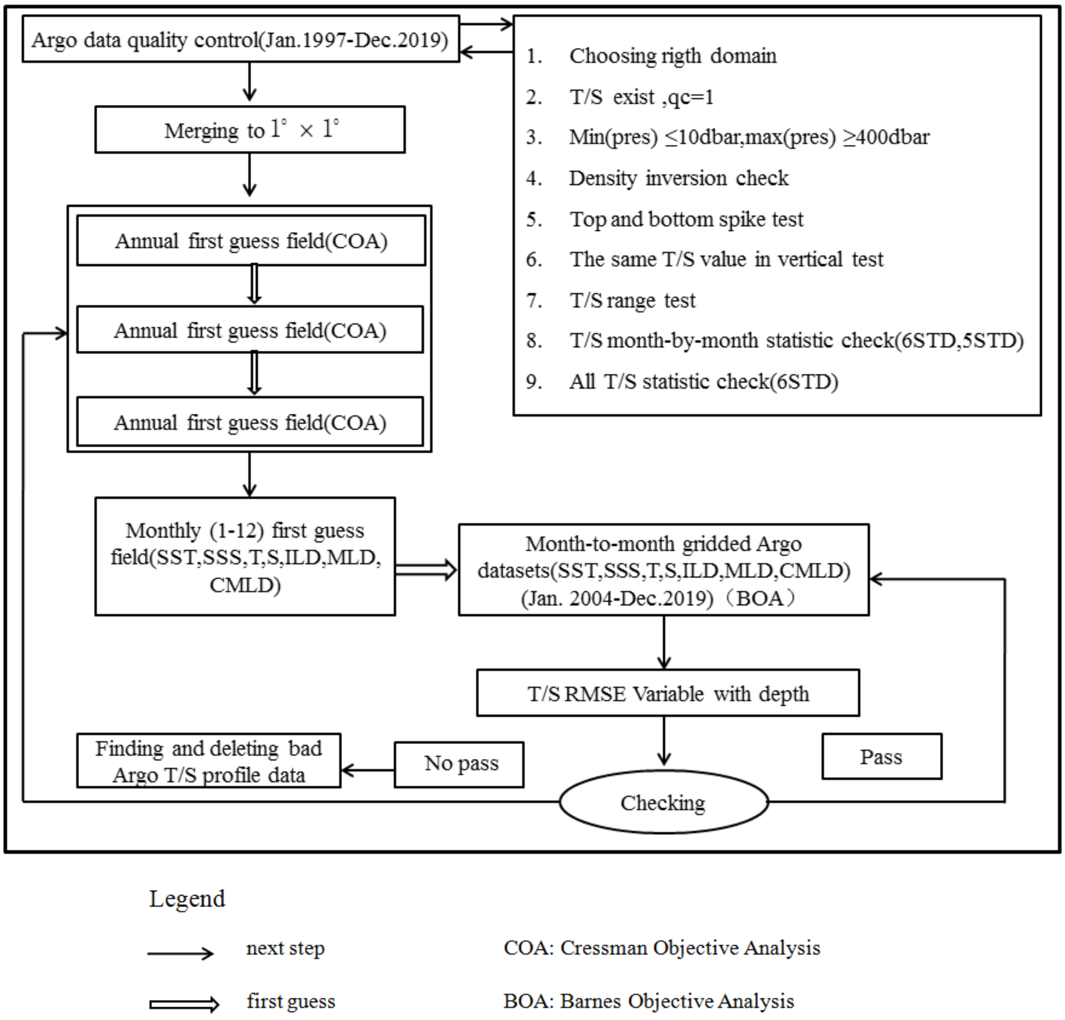

Figure 1 Scheme describing the production of the BOA_Argo gridded data set, including quality control and preprocessing of Argo data, as well as iterative applications of the objective analysis and RMSE check steps.

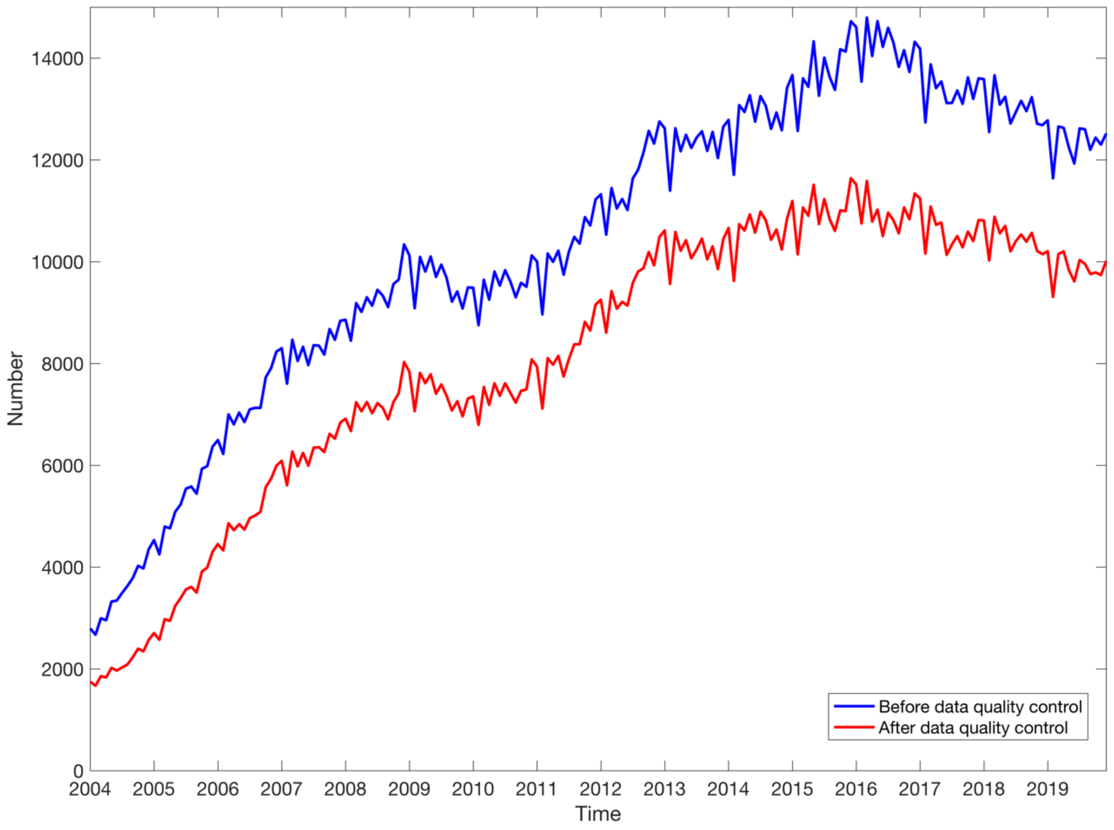

Figure 2 Number of Argo profiles before (blue) and after (red) a data screening from January 2004 to December 2019.

$$\begin{cases} w_{ib}^{n} = \dfrac{R_n^2 - r_{ib}^2}{R_n^2 + r_{ib}^2}, r_{ib}^2 < R_n^2 \\ w_{ib}^{n} = 0, r_{ib}^2 \geq R_n^2 \end{cases}, \qquad (2)$$

where r_{ib}^2 is the square of the distance between the observation *b* and the grid point *i*. R_n is usually taken to be several times the grid distance. We use a Cressman scheme (Eqns. 1&2) with three iterations to generate an annual mean climatology that serves as a first guess for the seasonal climatology. The scanning radii are set to **999 km, 666 km and 333 km**, respectively. Four seasonal means are then estimated using the same three radii of influence, with the annual mean as the initial condition. Finally, monthly mean climatologies are generated based on the same approach, with the corresponding seasonal mean as the first guess.

(3) Generation of gridded data using Barnes successive correction method

Monthly data are produced using Barnes successive correction method (*Barnes*, 1973), which is similar to the Cressman scheme in many aspects. *Barnes* (1973) defined the method by proposing a modified weighting function,

$$\begin{cases} w_{ib}^{n} = \exp(\frac{-r_{ib}^{2}}{\alpha\gamma}), r_{ib}^{2} < R_{n}^{2} \\ w_{ib}^{n} = 0, r_{ib}^{2} \geq R_{n}^{2} \end{cases}, \qquad (3)$$

where a is a filtering constant and g is the convergence factor. This modified weighting function reduces the number of iterations typically required to reach a stable solution. BOA_Argo applied the iterations twice, and the corresponding influence radiuses are selected both as 555 km, with the filtering parameters of 8.0×10^4 and 1.6×10^4 km^2, and the convergence factors of 0.2. After each iteration, a 9-point smoothing is applied twice. Consequently, BOA_Argo can retain larger portions of the mesoscale and large-scale signals for wavelength⩾5ΔX[2] (Li et al., 2017a).

(4) RMSEs of gridded T/S profiles

T/S profiles with large T/S RMSE values are removed, then the third (Barnes successive correction) and fourth (calculation of RMSEs) steps are repeated until all T/S RMSEs in the deep waters (>1500 dbar) are smaller than a threshold (RMSEs<0.06 ℃ for T，and RMSEs<0.01 for S). Temperature and salinity in the deep waters (>1500 dbar) may have a quite slow variability, so that T/S RMSEs at these depths should be small and consistent across months. Additional details are provided below.

Figure 3 shows the final global averaged T/S RMSEs at different depths. For water depths deeper than 1000 dbar, the average RMSE of temperatures is approximately 0.058±0.008 ℃, and the average RMSE of salinity is 0.008±0.001. Above 1000 dbar, the average RMSEs increase to 0.317±0.036 ℃ for temperature and 0.042±0.007 for salinity, probably due to the impacts of surface winds, heat fluxes, freshwater fluxes, and so on.

---

[2] For ΔX = 111 km, the meridional separation at the Equator

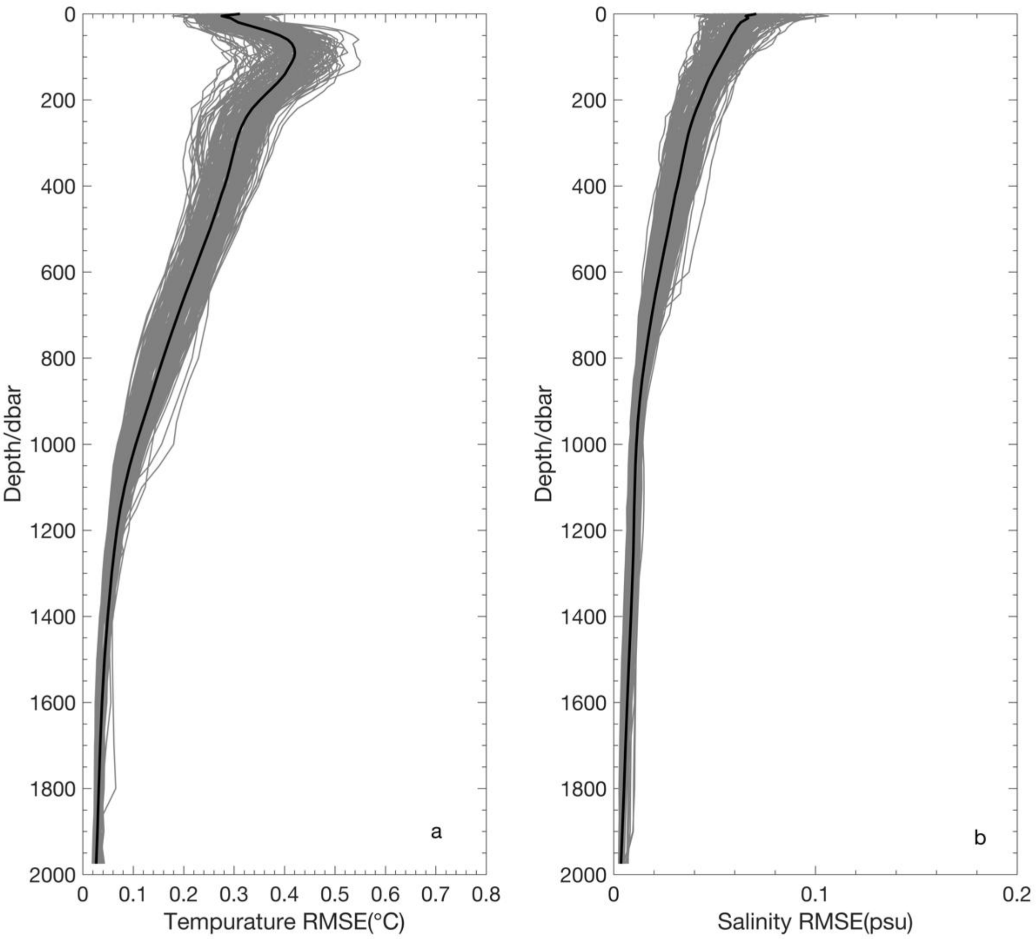

Figure 3 RMSEs of (a) temperature and (b) salinity between the gridded BOA_Argo analysis and merged observations after the objective analysis step (January 2004–December 2019). Solid black line denotes the average value of RMSEs for each variable, while solid grey lines denote the monthly RMSEs over the global oceans.

5. Data set description

Name: BOA_Argo.

Temporal coverage: From January 2004 to December 2019.

Temporal resolution: Monthly.

Spatial resolution: horizontal $1^o\times1^o$ (Longitude: 0.5:1.0:359.5, Latitude: -79.5:1.0:79.5); 58 levels in vertical from surface to 1975 dbar depth.

The MATLAB and NetCDF format versions are available.

5.1 MATLAB Version

For example:

BOA_Argo_2004_01.mat is the data file for January 2004, variables included:
lon (longitude, 360×160),
lat (latitude, 360×160),
pres (pressure,58, unit: dbar),
temp (temperature, 360×160×58, unit: ℃),
salt (salinity, 360×160×58, unit: PSS-78),
mld_t (Isothermal Layer Depth, 360×160, unit: m),
mld_dens (Mixed Layer Depth, 360×160, unit: m),
mld_composed(Composed Mixed Layer Depth, 360×160, unit: m).

BOA_Argo_annual.mat is the data file for annual climatology, variables included:
temp_annual (temperature, 360×160×58),
salt_annual (salinity, 360×160×58),
mld_t_annual (ILD, 360×160),
mld_dens_annual (MLD, 360×160),
mld_composed_annual (CMLD, 360×160).

BOA_Argo_monthly_1.mat is the data file for January climatology, variables included:
temp_monthly (temperature, 360×160×58),
salt_monthly (salinity, 360×160×58),
mld_t_monthly (ILD, 360×160),
mld_dens_monthly (MLD, 360×160),
mld_composed_annual（CMLD, 360×160）;
landmask.mat is the land mask file, variables included:
lons (longitude, 360×160),
lats (latitude, 360×160),

landmask (360×160×58), and land is flag 1,ocean is flag 0.

5.2.NetCDF format

‘**BOA_Argo_YYYY_MM.nc**’ is the NetCDF format data, for example:

BOA_Argo_2004_01.nc is the data file for January 2004, variables included:

lon (longitude, 360),

lat (latitude, 160),

pres (pressure, 58),

time (days since 0000-01-01, 1)

temp (temperature, 1×58×160×360),

salt (salinity, 1×58×160×360),

ILD (Isothermal Layer Depth, 1×160×360),

MLD (Mixed Layer Depth, 1×160×360),

CMLD (Composed Mixed Layer Depth, 1×160×360).

FillValue is 99999.

Below is a MATLAB script to read all the variables:

```
ncid=netcdf.open('BOA_Argo_2004_01.nc','nowrite');% read the file

varid=netcdf.inqVarID(ncid,'lat'); % get the latitude variable id
lat=netcdf.getVar(ncid,varid); % get the latitude value

varid=netcdf.inqVarID(ncid,'lon'); % get the longitude variable id
lon=netcdf.getVar(ncid,varid); % get the longitude value

varid=netcdf.inqVarID(ncid,'pres'); % get the pressure variable id
pres=netcdf.getVar(ncid,varid); % get the pressure value

varid=netcdf.inqVarID(ncid,'time'); % get the time variable id
time=netcdf.getVar(ncid,varid); % get the time value

varid=netcdf.inqVarID(ncid,'ILD'); % get the Isothermal Layer Depth(ILD) variable id
mld_t=netcdf.getVar(ncid,varid); % get the ILD value

varid=netcdf.inqVarID(ncid,'MLD'); % get the Mixed Layer Depth(MLD) variable id
mld_dens=netcdf.getVar(ncid,varid); % get the MLD value
```

```
varid=netcdf.inqVarID(ncid,'CMLD'); % get the Composed Mixed Layer
                                          Depth(CMLD)variable id
mld_cmd=netcdf.getVar(ncid,varid); % get the CMLD value

varid=netcdf.inqVarID(ncid,'temp'); % get the temperature variable id
temp=netcdf.getVar(ncid,varid); % get the temperature value

varid=netcdf.inqVarID(ncid,'salt'); % get the salinity variable id
salt=netcdf.getVar(ncid,varid); % get the salinity value

netcdf.close(ncid);
```

6. Acknowledgments

This work was supported by the Science and Technology Basic Work of the Ministry of Science and Technology of China (Grant No. 2012FY112300), the operation funds for basic scientific research of Second Institute of Oceanography, Ministry of Natural Resources (Grant No. JG1622), the National Natural Science Foundation of China (Grant Nos 41621064, 41406022, 41606003,U1811464)，and the Project of State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography (No.QNHX1916 and No.SOEDZZ1906).

7. References

Annie Wong, Robert Keeley, Thierry Carval and the Argo Data Management Team (2015), Argo Quality Control Manual For CTD and Trajectory Data, http://dx.doi.org/10.13155/33951.

Barnes, S. L (1964), A technique for maximizing details in numerical weather map analysis, *J. Appl. Meteor.*,3: 396–409.

Barnes, S. L. (1973), Mesoscale objective analysis using weighted time series observations. NOAA Tech. Memo. ERL NSSL-62, National Severe Storms Laboratory, Norman, OK, 41pp.

Chu P C, Fan C W, 2011. Maximum angle method for determining mixed layer depth from sea glider data. J Oceanogr, 67: 219-230

Good, S. A., M. J. Martin, N. A. Rayner (2013), EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res.,* 118,6704-6716, doi:10.1002/2013JC009067.

Hosoda, S., T. Ohira, T. Nakamura (2008), A monthly mean data set of global oceanic temperature and salinity derived from Argo float Observations. *JAMSTEC Rep. Res. Dev.,*Volume 8, November 2008,47-59.

Levitus, S. (1982), Climatological atlas of the World Ocean. NOAA Prof. Pap. 13, U.S. Gov. Print.

Off.,Washington, D. C.

Li Hong, Xu Jianping, Liu Zenghong, et al. (2012), Global ocean Argo gridded data set (BOA_Argo) user manual. Hangzhou: China Argo Real-time Data Center, 1-31 (in Chinese).

Li Hong, Xu Jianping, Liu Zenghong, et al. (2013), Study on the global ocean Argo gridded data set and its validation. Marine Science Bulletin, 32(6): 108-118 (in Chinese).

Li Hong, Xu Jianping, Liu Zenghong, et al. (2014), Global ocean Argo gridded data set (BOA_Argo) user manual. Hangzhou: China Argo Real-time Data Center, 1-31 (in Chinese).

Li Hong, Liu Zenghong, Xu Jianping, et al. (2015), Introduction of global ocean Argo gridded data set (BOA_Argo). Hangzhou: China Argo Real-time Data Center, 1-97 (in Chinese).

Li Hong, Liu Zenghong, Xu Jianping, et al. (2016), Introduction of global ocean Argo gridded data set (BOA_Argo). Hangzhou: China Argo Real-time Data Center, 1-190 (in Chinese).

Li, H., F. Xu, W. Zhou, D. Wang,J. S. Wright, Z. Liu, and Y. Lin (2017a), Development of a global gridded Argo data set with Barnes successive corrections. *J. Geophys. Res. Oceans*, 122, doi:10.1002/2016JC012285.

Li Hong, Liu Zenghong, Xu Jianping, Wu Xiaofen, Sun Chaohui, Lu Shaolei, Cao Minjie.(2017b). Manual of Global Ocean Argo gridded data set (BOA_Argo) (Version 2017), 13 pp.

Locarnini, R. A., A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, C. R. Paver, J. R. Reagan, D. R. Johnson, M. Hamilton, D. Seidov (2018), World Ocean Atlas 2013, Volume 1: Temperature. S. Levitus, Ed.; A. Mishonov, Technical Ed.; NOAA Atlas NESDIS 73, 40 pp.

Roemmich, D., and J. Gilson (2009), The 2004-2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from Argo program. *Progr. Oceanogr.* , 82:81-100.

Troupin, C., F. MachF. MachOuberdous, D. Sirjacobs, A. Barth, and J. ‑ M. Beckers (2010), High ‑ resolution climatology of the northeast Atlantic using Data ‑ Interpolating Variational Analysis (Diva). *J. Geophys. Res.*, 115C08005,doi:10.1029/2009JC005512.

8. Contacts

Any comments, questions regarding this Argo gridded data set can be directed to the co-authors:

Lu Shaolei	E-mail:	lsl324004@163.com
Li Hong	E-mail:	slvester_hong@163.com
Liu Zenghong	E-mail:	liuzenghong@139.com

9. Copyrights and conditions

The BOA_Argo gridded data set generated by China Argo Real-time Data Center (CARDC, http://www.argo.org.cn/) is open for an unrestricted usage, any copying and distribution of this data set are permitted.

Before using the data, please read the conditions below and acknowledge your acceptance.

Conditions:

The user acknowledges that the Argo data product was developed by CARDC for research purposes. The CARDC will not be liable for interpretation of or inconsistencies, discrepancies, errors or omissions in any or all of the product as supplied.

CARDC shall not be liable for any loss or damage when the user makes use of this data set.

The user agrees that whenever the data set is used in publications, the CARDC shall be acknowledged as the source of the product, and with the following form:" The BOA_Argo dataset used in this study is produced at China Argo Real-time Data Center".