Manual of Global Ocean Argo Gridded Dataset (GDCSM_Argo) (Version 2022)

Chunling Zhang[1]，Danyang Wang[1], Shaolei Lu[2]，Zenghong Liu[2]，Zhaoqin Li[2]，Xiaofen Wu[2]，Jianping Xu[2], Chaohui Sun[2]

1. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China
2. State Key Laboratory of Satellite Ocean Environment Dynamics (SOED), Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China

Shanghai, China

August 2022

This document should be cited as:

Chunling Zhang, Danyang Wang, Shaolei Lu，Zenghong Liu，Zhaoqin Li，Xiaofen Wu， Jianping Xu, Chaohui Sun (2022). Manual of Global Ocean Argo gridded dataset (GDCSM_Argo) (Version 2022), 13 pp.

For details about the development of GDCSM_Argo data set, please refer to the paper " Zhang, C.; Wang, D.; Liu, Z.; Lu, S.; Sun, C.; Wei, Y.; Zhang, M. Global Gridded Argo Dataset Based on Gradient-Dependent Optimal Interpolation. J. Mar. Sci. Eng. 2022,10,650. https://doi.org/ 10.3390/jmse10050650 ".

Contents

1. History

The development of GDCSM_Argo gridded dataset includes three versions of improvement (Table 1). The first version developed in September 2013 is mainly devoted to the verification of gradient-dependent correlation scale method (GDCSM, also named gradient-dependent optimal interpolation). The gridded dataset of the first version only covered the Pacific Ocean and contained 26 vertical levels (0-2000 m) for less calculation (Zhang et al., 2013). Then we improved the background and optimized the algorithm of correlation scale. A global ocean Argo gridded dataset, the second version of GDCSM_Argo, was developed (Xie et al., 2019) in May 2019. This gridded dataset is the third version of GDCSM_Argo. It provides more accurate multi parameter analysis results with higher vertical resolution by improving the parameters setting in the objective analysis system (Zhang et al., 2022a).

Table 1. History of the GDCSM_Argo gridded dataset

Version	**GDCSM_Argo 2013**	**GDCSM_Argo 2019**	**GDCSM_Argo 2022**
Domain	Pacific Ocean (120°E~70°W,60°S~60°N)	Global Ocean (180°E~180°W,60°S~60°N)	Global Ocean (179.5°E~179.5°W,89.5°S~89.5°N)
Data source	Argo T/S	Argo T/S	Argo T/S
First guess	Cressman successive correction, Argo 2004-2011 T/S climatology data	Optimal interpolation, Argo 2004-2017 T/S climatology data	Gradient-dependent optimal interpolation, Argo 2004-2021 T/S climatology data
Objective analysis method	Gradient-dependent correlation scale method	Gradient-dependent correlation scale method	Gradient-dependent optimal interpolation
Horizontal resolution	1 degree	1 degree	1 degree
Vertical resolution	26 levels, 0-1975 dbar	26 levels, 0-1975 dbar	58 levels, 0-1975 dbar
Temporal coverage, resolution	2004-2011, monthly	2004-2017, monthly	2004-2021, monthly

Variable	T/S	T/S	T/S/ SV/MLD/TBD/TTG[1]

2. Introduction

It's well known that the observation profiles collected by the autonomous profiling floats are randomly distributed in space and time, and significantly limits the use of these data, especially in operational applications. The optimal interpolation or variational analysis is proven to be an effective method to construct the gridded products. However, determining an appropriate scale for spatial correlation is a key problem in optimal interpolation or variational analysis. Many studies have calculated the deviation between observational information and background information to obtain the variance in the background error and the scale of correlation through function fitting (Bonekamp et al.,2001; Hollingsworth et al., 1986; Meyers et al., 1991). This requires sufficiently dense observational data to provide multiple scales of oceanic information. Moreover, the estimated correlation scales are usually constant. However, the scales of spatial correlation generally vary with the factors used for analysis, direction, and location. To solve this problem, Zhang et al. (2013) developed a gradient-dependent correlation scale method (GDCSM, also called gradient-dependent optimal interpolation) based on horizontal changes in the factors of analysis. The thermohaline structures of the ocean can be described more accurately by a scheme based on an anisotropic correlation scale because it is more flexible and applicable to multi-factor analysis (Zhang et al., 2021; Zhang et al., 2022b). Gradient-dependent optimal interpolation provides an important approach for constructing more accurate gridded Argo products.

Another crucial problem is that most Argo profiles lack sea surface information at present. Taking the shallowest observation depth as the sea surface is one solution to this problem that does not require adding other observations (Yan et al., 2010). For most Argo profiles, the differences between the shallowest measurements and those at sea surface cannot negligible. Another solution is to merge surface information from traditional observations (e.g., XBT, CTD, and TAO) (Hosoda et al., 2008; Martin et al., 2007). However, there are far fewer traditional sea surface observations than that of Argo profiles. Besides, these sea surface observations are expected to be mismatched with Argo observations

[1] **T:** temperature, **S:** salinity, **SV:** sound velocity, **MLD:** Mixed Layer Depth, **TBD:** Thermocline Bottom Depth, **TTG:** Thermocline Temperature Gradient.

in terms of quantity and time for the near future. The inversion of surface data corresponding to Argo profiles by using statistical models is thus important.

Then a global Argo gridded dataset, named GDCSM_Argo, is developed by an objective analysis system based on gradient-dependent optimal interpolation method and a pycnocline-based model.

3. Scheme of production of the GDCSM_Argo gridded dataset

An objective analysis system based on gradient-dependent optimal interpolation was used to build a global Argo gridded dataset, called GDCSM_Argo, by using the procedure illustrated in Figure 1. First, the density, sound velocity, and parameters of the thermocline corresponding to each Argo profile were calculated and screened. Second, the gradient-dependent scales of correlation were given by setting the background data via the procedures detailed in Section 3.3. Third, objective analysis of the Argo data was carried out based on gradient-dependent optimal interpolation and the global subsurface gridded data of the temperature, salinity, sound velocity, and the gridded parameters of the thermocline was obtained. Finally, the pycnocline-based model was used to construct the surface temperature and salinity. The gridded sound velocity on the surface was determined by the surface temperature and salinity.

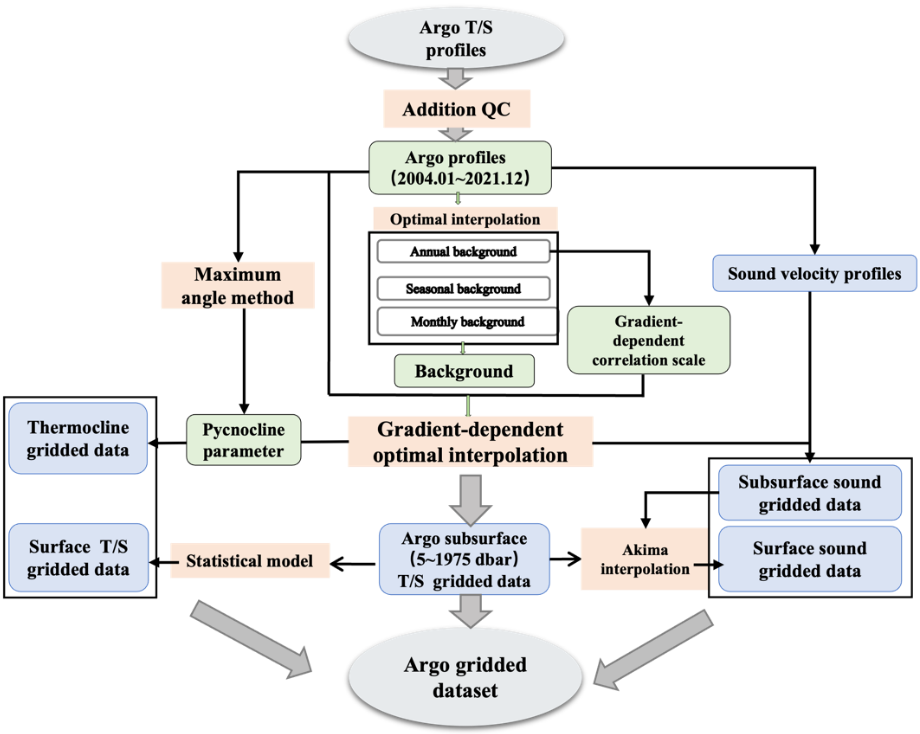

Figure 1. Flowchart describing the generation of the GDCSM_Argo gridded dataset based on gradient-dependent optimal interpolation.

3.1 Data preparing

The Argo T/S profiles were from the China Argo Real-Time Data Center (CARDC, ftp://ftp.argo.org.cn/pub/ARGO/global/). The Argo float network has provided global coverage since 2004. We collected Argo observations from 1 January 2004 to 31 December 2021 to generate a monthly gridded dataset. A total of 2,357,185 T/S profiles were retained after a post-quality-control procedure developed by CARDC (Hosoda et al., 2008; Li et al., 2020). These profiles were interpolated to 57 vertical levels (5 dbar, 10–200 dbar at 10 dbar intervals, 220–500 dbar at 20 dbar intervals, 550–1250 dbar at 50 dbar intervals, 1300–1900 dbar at 100 dbar intervals, and 1950 dbar) by using Akima interpolation (Akima et al., 1970). The profiles of sound velocity were calculated via the T/S profiles by using the formula for sound velocity (Fofonoff et al., 1983). And the parameters of the thermocline (mixed layer depth, MLD; thermocline bottom depth, TBD; and thermocline temperature gradient, TTG) are obtained by the maximum angle method (Chu et al., 2011).

3.2 Background fields

The gridded dataset developed in this study was based only on Argo observation data, without any other observation or output of numerical simulation. The arithmetic average and traditional optimal interpolation (Riishogaard et al., 1998; Gandin et al., 1963) were used to construct the background of the subsurface (5–2000 dbar) temperature, salinity, and sound velocity from Argo. The Argo profiles were first filtered to remove unreliable data and merged into $1^{\circ}\times1^{\circ}$ boxes in the global ocean (179.50°W–179.50°E, 89.5°S–89.5°N). Based on this merged fields, the climatological background fields were constructed by traditional optimal interpolation. Then, the seasonal background fields were constructed by taking climatological background field as the initial fields via the gradient-dependent optimal interpolation. Finally, the monthly background fields were derived from the seasonal average fields.

3.3 Generation of gridded dataset using Gradient-Dependent Optimal Interpolation

Monthly data are produced using Gradient-Dependent Optimal Interpolation. The results at the gridded point used for analysis were adopted as the background value plus the observational increments weighted by optimal weights. The standard equation of the influence of M observations on the analysis point is given in Equation (1). A critical part of the scheme involves estimating the optimal weights. According to minimum variance theory, the optimal weight can be determined by solving Equation (2) (Gandin et al., 1963) (pp. 168–203):

$$v_i^a = v_i^b + \sum_j^M w_{ij}\,\delta y_j^o \tag{1}$$

$$\sum_{j=1}^M w_{ij}\,\mu_{jk} + \eta_k w_{ik} = \mu_{ik}, \qquad k = 1,\cdots,M \tag{2}$$

where v_i^a is the analysis value and v can be any environmental variable, such as temperature, salinity, or sound velocity. The symbol v_i^b, given by the Argo monthly background, is the first estimated value. The subscript i denotes the number of gridded points used for analysis, and j and k denote the number of available sites of Argo profiles. For observational increments, $\delta y_j^0 = y_j^0 - H(v_j^b)$, the observational operator H was used to convert the background into the first guesses of the observation y_j^0. A radial distance was set to ensure that only Argo profiles were located within a specified range surrounding each analysis point. Each increment had an optimal weight w_{ik} associated with the background error correlations μ_{jk} and μ_{ik}. μ_{jk} and μ_{ik} are correlations between the background errors at the two observational points, j and k, and at the gridded and observation points i and k, respectively. The parameter η_k is the square of relative observational errors compared with the background errors. It is frequently set to be a constant η for a single source of observation and "tuned" to vary the weights of the observations. The root mean-squared error (RMSE) for each factor was at a minimum with η = 0.5 in the global ocean (Zhang et al., 2022a).

The correlations are usually assumed to follow a Gaussian exponential function, and are inversely proportional to the distance (Zhang et al., 2013; Zhang et al., 2021; Kalnay et al., 2003), as shown in Equations (3) and (4):

$$\mu_{ik} \sim exp\left[-\frac{(x_i-x_k)^2}{\left(L_\emptyset/G_x\right)^2} - \frac{(y_i-y_k)^2}{\left(L_\emptyset/G_y\right)^2}\right] \tag{3}$$

$$G_x = 1 + \frac{|\partial v/\partial x|}{E(|\partial v/\partial x|)}, \qquad G_y = 1 + \frac{|\partial v/\partial y|}{E(|\partial v/\partial y|)} \tag{4}$$

where x and y are the longitude and latitude, respectively, and $L_\emptyset/G$ depends on the Rossby radius of deformation and changes in the horizontal gradient that define the scale of correlation of the background error. The parameter $L_{\emptyset}$ is the scale of correlation that can be obtained from the product of the scale parameter and the cosine function of the latitude $\emptyset$ at the analysis gridded point. The radial distances were set to 500 km and 1000 km when constructing the climatological and the monthly products, respectively, to ensure that a sufficient number of observations were considered (Zhang et al., 2021; Zhang et al., 2022a). The parameter G, calculated using Argo climatological data, was associated with the horizontal gradients at location i. It contained a zonal component G_x and a meridional component G_y.

Taking temperature as an example, Figure 2 shows the zonal and meridional distributions of the correlation scale at a depth of 100 dbar, with each small ellipse generated in $3^{\circ}\times3^{\circ}$ boxes. The scales of zonal correlation were different from the meridional scales at each gridded point used for analysis. The correlation scales varied with the horizontal gradient of temperature. In particular in areas of large temperature gradient, such as the Kuroshio, the Gulf Stream, and the subtropical composite area, the scale of temperature correlation was relatively small. It is clear that the Gaussian functions presented in Equations (3) and (4) provide error correlations in the anisotropic background at each point of analysis.

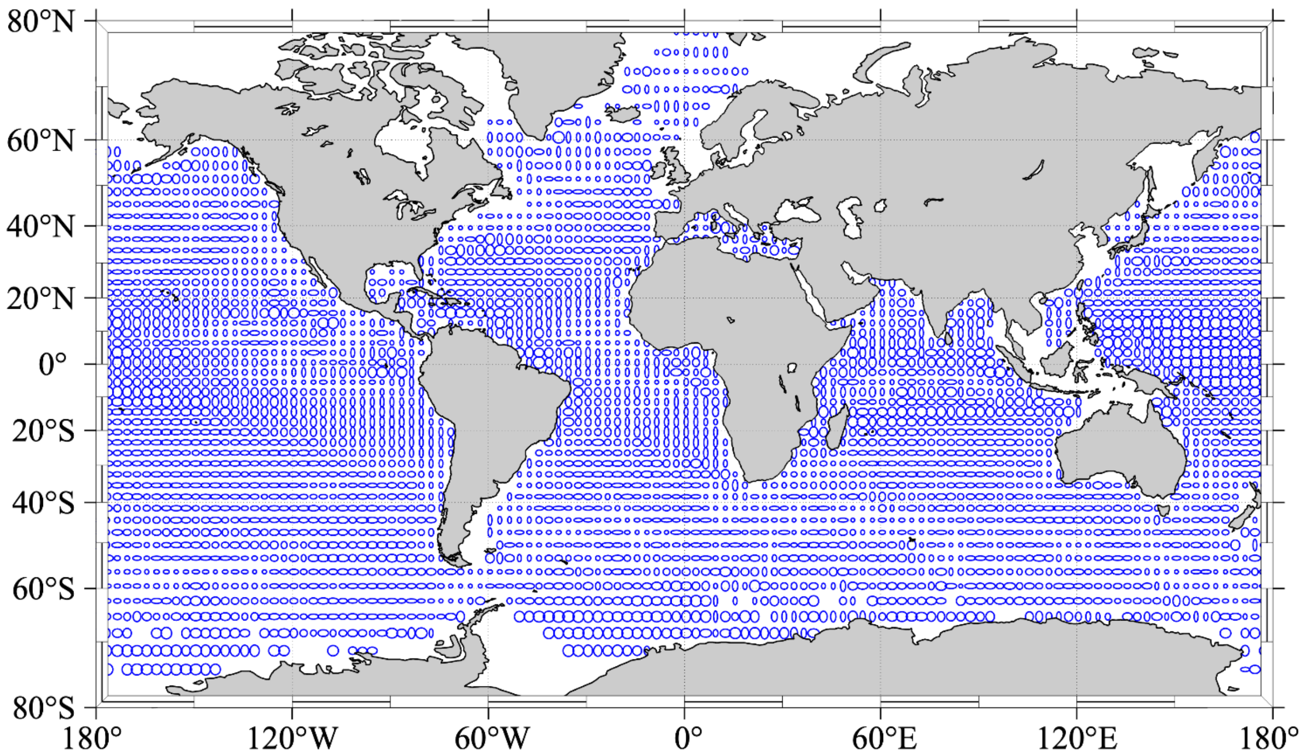

Figure 2. The distribution of the temperature correlation scales at 100 dbar.

3.4 Inversion of surface information by a pycnocline-based model

The statistical model used to estimate the surface temperature and salinity was based on the following parameters: the MLD or upper depth of the thermocline, TBD, and TTG (Chu et al., 2011). Density profiles were used to calculate the upper depth and bottom depth of the pycnocline, denoted by MLD and TBD, respectively. It is more reasonable to obtain the parameters of the thermocline by density than temperature to avoid the influence of the barrier layers and salinity (Zhang et al., 2015). These key parameters were determined by the maximum angle method (Zhang et al., 2022c; Wang et al., 2022), and the surface information was constructed by the model depicted in Equations (5)–(7):

$$SST = \frac{P_0}{P_z} T_z \quad , \; SSS = S_z \; , \; (0 \le z \le MLD) \tag{5}$$

$$SST = T_z - TTG(z - MLD) \quad (MLD \le z \le TBD) \tag{6}$$

$$SST_i = \frac{\sum_{j=1}^{9} b_{i,j} SST_{i,j}}{\sum_{j=1}^{9} b_{i,j}}, \qquad b_{i,j} = exp\left(-\left(r_{i,j} - \bar{r}_i\right)^2 / L^2\right) \tag{7}$$

where SST and SSS represent the sea surface temperature and salinity, respectively, T_z and S_z indicate the subsurface temperature and salinity at the depth z as estimated by data on the density profile from Argo, respectively, and P_0 and P_z are pressures on the sea surface and the subsurface, respectively. SST is significantly affected by the characteristics of the thermocline. Therefore, nine reference layers (5, 10, 20, 30, 40, 50, 80, 100, and 120 dbar) were selected to calculate the sea surface temperatures by Equation (5) and (6), and were mean-weighted by Equation (7) (Zhang et al., 2015). The subscripts i and j denote the number of Argo profiles available around the gridded point for analysis and the datum layer. $b_{i,j}$ is the weight coefficient and $r_{i,j}$, with the mean value $\bar{r}_i$, represents the RMSE of $SST_{i,j}$ compared with the value obtained by the Global Temperature and Salinity Profile Programme (GTSPP). The parameter L of the correlation scale was set to 2 degrees, as in previous studies (Zhang et al., 2015).

4. Data set description

Name: GDCSM_Argo.

Temporal coverage: From January 2004 to December 2021.

Temporal resolution: Monthly.

Spatial resolution: horizontal $1^{o}\times1^{o}$ (Longitude: -179.5:1.0:179.5, Latitude: -89.5:1.0:89.5); 58 levels in vertical from surface to 1975 dbar depth.

The MATLAB and NetCDF format versions are available.

4.1 MATLAB Version

For example:

GDCSM_Argo_annual.mat is the data file for annual climatology,

GDCSM_Argo_month_*.mat is the data file for monthly climatology,

GDCSM_Argo_yyyymm.mat is the data file for mm/yyyy (mm: month; yyyy: year),

variables included:

lon (longitude, 360×180),

lat (latitude, 360×180),

pres (pressure,58, unit: dbar),

time (time, 1),

gdcsm_temp (temperature, 360×180×58, unit: ℃),

gdcsm_psal (salinity, 360×180×58, unit: PSS-78),

gdcsm_svel (sound velicity, 360×180×58, unit: m/s),

gdcsm_mld (Mixed Layer Depth, 360×180, unit: dbar),

gdcsm_tbd (Thermocline Bottom Depth, 360×180, unit: dbar),

gdcsm_ttg (Thermocline Temperature Gradient, 360×180, unit: ℃/dbar).

4.2 NetCDF format

'GDCSM_Argo_annual.nc', 'GDCSM_Argo_month_*.nc', 'GDCSM_Argo_yyyymm.nc' are the NetCDF format data respectively, for example:

GDCSM_Argo_200401.nc is the data file for January 2004,

variables included:

lon (longitude, 360),

lat (latitude, 180),

pres (pressure, 58),

time (time,1),

temp (temperature, 1×58×180×360),

salt (salinity, 1×58×180×360),

svel (sound velocity, 1×58×180×360),

MLD (Mixed Layer Depth, 1×180×360),

TBD (Thermocline Bottom Depth, 1×180×360),

TTG (Thermocline Temperature Gradient, 1×180×360).

FillValue is 99999.

5. Acknowledgements

The Argo profiles were collected and made freely available by the International Argo Program and national programs that contribute to it. The Argo Program is part of the Global Ocean Observing System. We are grateful to the China Argo Real-Time Data Center for the post-quality-controlled Argo dataset, the Argo Program Office for distributing the gridded Argo datasets, and the National Oceanographic Data Center for providing us with the climate data. We thank Dr. H.L. from Fudan University for his technological assistance during the preparation of this manuscript.

6. References

Akima, H.1970. A new method for interpolation and smooth curve fitting based on local procedures. Journal of the Acm, 17(4):589~602.

Bonekamp, H.G., Oldenborgh, J.V., Burgers, G. 2001. Variational assimilation of TAO and XBT data in the HOPE OGCM: Adjusting the surface fluxes in the tropical ocean. Geophysical Research, 106:16693~16709.

Chu, P.C., Fan, C. 2011. Maximum angle method for determining mixed layer depth from sea glider data. Journal of Oceanography, 67(2):219~230.

Fofonoff, N.P., Millard, R.C. 1983. Algorithms for computation of fundamental properties of seawater. Unesco Technical Papers in Marine Science 44; UNESCO: Paris, France, pp. 3~59.

Gandin, L.S. 1963. Objective analysis of meteorological fields. Leningrad: Gidromet, pp. 242.

Hollingsworth, A., Lonnberg, P. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. Tellus Series A-dynamic Meteorology & Oceanography, 38A (2):111~136.

Hosoda, S., Ohira, T., Nakamura, T. 2008. A monthly mean dataset of global oceanic temperature and salinity derived from Argo float observations. JAMSTEC Report of Research and Development, 47~59.

Kalnay, E. 2003. Atmospheric modeling, data assimilation, and predictability. Cambridge University Press: London, UK, pp. 129.

Li Zhaoqin, Liu Zenghong, Lu Shaolei. 2020. Global Argo data fast receiving and post-quality-control-system. In IOP Conference Series: Earth and Environmental Science; IOP Publishing: Bristol, UK.

Martin, M.J., Hines, A., Bell, M.J. 2007. Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact. Quart. J. Roy. Meteorol. Soc, 133:981~995.

Meyers, G., Phillips, H., Smith, N., et al. 1991. Space and time scales for optimal interpolation of temperature-Tropical Pacific Ocean. Progress in Oceanography, 28:189~218.

Riishogaard, L.P. 1998. A direct way of specifying flow-dependent background error correlations for meteorological analysis system. Tellus A, 50: 42~57.

Wang, X., Cheng, L., Zhang, C., Li H. 2022. Comparative Analysis of Global Applicability of Different Mixed Layer Depth Algorithms. HaiYang TongBao, accepted.

Xie, C., Xu, M., Cao, S., Zhang, Y., Zhang, C. 2019. Gridded Argo data set based on GDCSM analysis technique: establishment and preliminary application. Journal of Marine Science, 37(4):24~35.

Yan, C., Zhu, J., Xie, J. 2010. An ocean reanalysis system for the joining area of Asia and Indian-Pacific Ocean. Atmos. Oceanic Sci. Lett., 3:81~86.

Zhang, C., Wang, D., Liu, Z., Lu, S.; Sun, C.; Wei, Y.; Zhang, M. 2022a. Global Gridded Argo Dataset Based on Gradient-Dependent Optimal Interpolation. J. Mar. Sci. Eng. 10,650. https://doi.org/10.3390/jmse10050650.

Zhang, C., Wang, D., Wang, Z. 2022b. Fishery analysis using gradient-dependent optimal interpolation. Acta Oceanologica Sinica, 41(2):116~126.

Zhang, C., Zhang, M., Wang, Z., Yang, S. 2022c. Thermocline model for estimating Argo surface temperature. Sustainable Marine Structure, 4(1): 1~12.

Zhang, C., Wang, Z., Liu, Y. 2021. An Argo-based experiment providing near-real-time subsurface oceanic environmental information for fishery data. Fisheries Oceanography, 30: 85~98.

Zhang, C., Xu, J., Bao, X., Wang, Z. 2013. An Effective Method for Improving the Accuracy of Argo

Objective Analysis. Acta Oceanololgica Sinica, 32(7): 66~77.

Zhang, C., Xu, J., Bao, X. 2015. Gradient dependent correlation scale method based on Argo. Journal of Army Engineering University Science Technology, 16(5): 476~483.

7. Contacts

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

Zhang Chunling E-mail: clzhang@shou.edu.cn

Lu Shaolei E-mail: lusl@sio.org.cn

Liu Zenghong E-mail: zliu@sio.org.cn

8. Copyrights and conditions

The GDCSM_Argo gridded dataset generated by Shanghai Ocean University (SHOU) and 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 SHOU and CARDC for research purposes. The SHOU and CARDC will not be liable for interpretation of or inconsistencies, discrepancies, errors or omissions in any or all of the product as supplied. SHOU and CARDC shall not be liable for any loss or damage when the user makes use of this dataset.

The user agrees that whenever the dataset is used in publications, the SHOU and CARDC should be acknowledged as the source of the product, and with the following form: "The GDCSM_Argo dataset used in this study is produced by Shanghai Ocean University and China Argo Real-time Data Center".