A new automatic quality control system for ocean profile observations and impact on ocean warming estimate

Zhetao Tan[1,6], Lijing Cheng[1,2,6*], Viktor Gouretski[1,3], Bin Zhang[2,4], Yanjun Wang[2,4], Fuchao Li [2,4], Zenghong Liu[5], Jiang Zhu[1]

[1] International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

[2] Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China

[3] Integrated Climate Data Center, Center for Earth System Research and Sustainability, the University of Hamburg, Germany

[4] Oceanographic Data Center, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

[5] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China

[6] University of Chinese Academy of Sciences, Beijing, China

Corresponding author: Lijing Cheng (chenglij@mail.iap.ac.cn)

Key Points

- A new climatological range-based automatic quality control algorithm for ocean temperature observations was developed and evaluated.
- The new algorithm is effective in removing outliers meanwhile minimizing the percentage of mistakenly discarding good data.
- A non-negligible impact of quality control in ocean heat content estimate is found.

Abstract

The rapidly enhancing global hydrographic in-situ observations archive is data quality heterogeneous. Different data applications (e.g., climate change science) require a high-performance quality control (QC) system to reliably identify outliers in profile data. This study presents a new automatic QC procedure (CAS-Ocean Data Center quality control system; CODC-QC) for ocean in-situ temperature outliers detection. Unlike many existing QC procedures, no assumption is made of a Gaussian distribution law in CODC-QC as the oceanic variables are typically skewed. Instead, the 0.5% and 99.5% quantiles are used to define the local temperature climatological ranges. Additionally, we constructed local climatological ranges for the vertical temperature gradient which increased the ability of identifying spurious profiles. The performance of CODC-QC was evaluated using two benchmark datasets. Results demonstrated that CODC-QC is effectively in removing spurious data and minimizing the percentage of mistakenly flagged good data. Additionally, the CODC-QC was applied to the global World Ocean Database (WOD) historical temperature profiles and a significant quality-dependent on instrumentation types was found. Finally, as ocean heat

content (OHC) is a fundamental indicator of climate change, the impact of different QC systems on OHC estimation is examined. Results based on an existing mapping approach indicate that applying CODC-QC system leads to a 41.7 % (4.9%) difference for linear trend of the global 0-2000m OHC changes within 1955-1990 (1991-2020) compared to the WOD-QC, implying a non-negligible source of error in ocean warming estimate. The new QC procedure could support further improvement of the oceanic climate records and other applications.

Plain Language Summary

The global archive of hydrographic in-situ observations is data quality heterogeneous, and thus it needs to apply the quality control (QC) to support the observation-based climate change science study (e.g., ocean warming monitoring). There are several existing QC algorithms to identify outliers, but most of them assume a Gaussian distribution law to the ocean variables, which is not an accurate estimation. In this study, we present a new climatological range-based automatic QC algorithm for ocean in-situ temperature observations. The new local climatological temperature and temperature gradient ranges fields without assuming any distribution laws are applied to this QC algorithm. Two benchmark datasets are used to assess the capability of effectively detecting outliers with minimizing the possibility of mistakenly discarding good data. Finally, as ocean heat content (OHC) is one of the most important indicators of climate change (especially for oceans warming), the new QC algorithm was applied to historical temperature datasets and calculated the OHC. A non-negligible impact of QC on OHC estimate is found. This study will promote the important role of QC for ocean and climate studies.

Keywords: Quality control, Temperature, Outlier, Climatology, Ocean heat content, Climate

1. Introduction

First released in 1994 (Levitus & Boyer, 1994) the digital archive of hydrographic data – the World Ocean Database (WOD) - now comprises about 20 million temperature profiles. There is an increasing demand for this data within a broad scientific community (Stammer et al., 2019), governmental and non-governmental organizations (IPCC, 2021), industry, fisheries (Brander, 2010) and individuals. The characteristic feature of this archive is its heterogeneity as it includes the data on temperature, salinity, and other oceanic parameters measured by more than a dozen of instrumentation types during the past ~150 years (Davis et al. 2019).

Instruments and methods used by physical oceanographers undergone significant changes during more than a century long history of observational oceanography. Therefore, measuring temperature by means of mercury-in-glass thermometers attached to the wire (old Nansen cast technique from 1890s) gave place to the electronic instruments like ship-born CTD (Conductivity-Temperature-Depth sonde) and XBT (eXpendable bathythermograph) during the late 1960s, followed by the development and implementation of autonomous Argo profiling

floats and gliders in this century (Garcia et al., 2018). Each instrumentation provides data of different accuracy and quality, which depend on the measurement devices, measurement techniques, operator experiences and weather conditions. Additionally, the accuracy of the data is also linked to the respective quality/accuracy requirements related to the scientific or other goals of the specific cruise or expedition. Whereas the accuracy of several tenth of degree might be sufficient for the detection of oceanic fronts or the location of thermocline, revealing secular temperature changes in the deep ocean requires the accuracy of thousandth of a degree.

The other characteristic feature of the global oceanographic archive is the incompleteness of the metadata which provide information on how the observations have been done. The absence of the metadata is an obstacle for the data-quality assessment and the development of correction schemes (see Cheng et al. (2014) and Gouretski and Cheng (2020) for metadata problems related to XBT and MBT profiles). Additionally, the hydrographic data are characterized by random and systematic errors specific for each instrumentation type. Therefore, quality control (QC) of observed data represents one of the most important tasks for the development of an oceanographic database (Domingues & Palmer, 2015).In early years, the data to be analyzed were typically limited to the observations obtained during one or several cruises, so that the QC was usually performed manually by the experts (Gronell & Wijffels, 2008). However, the manual QC of large regional or global datasets is not feasible because of the amount of data, so that the automated QC procedures (AutoQC) are required (Goni et al., 2019; Roemmich et al., 2019). The first digital atlas of the world ocean (Levitus, 1982) provides one of the first examples where the AutoQC was implemented to the global dataset. In recent years, some novel methods (e.g., machine learning) related to AutoQC have been developed (e.g., Mieruch et al. (2021); Ono et al. (2015); Sugiura and Hosoda (2020)).

The aim of AutoQC procedure is to identify and flag outliers, that is observations which differ significantly from the majority of other data points in the population (Tan et al., 2022). It should be noted that outliers not necessarily represent erroneous data and can simply reflect the natural variability of the measured variable. There is no rigid mathematical definition of an outlier. Therefore, any QC procedure inherently possess a certain degree of subjectivity as the outliers are defined by means of subjectively set thresholds (Argo, 2000).

Except for the surface measurements and the data from the underway thermographs, the global hydrographic archive essentially constitutes the archive of profile data. Typically, a hydrographic profile consists of the number of observations at different depth levels between the ocean surface and the ocean bottom. To increase the reliability in detection of erroneous data, a suite of quality checks is applied to each profile. The checks in AutoQC can be subdivided in several groups: 1) comparison of each observation with predefined local or global variable range; 2) checking profile shape; 3) comparing profile attributes with those for the respective instrumentation type. The larger the

number of distinct quality checks failed, the higher is the probability that the flagged observation really represent an outlier.

Until recently, the AutoQC typically assumes the Gaussian distribution law, with the outliers being defined as data points which deviate by more than predefined multiple times the standard deviations from the mean. For example, the WOD-QC procedure applies 3 to 5 standard deviation criteria in 5-degree boxes (Garcia et al., 2018). Since statistical distributions of temperature, salinity and other parameters in the ocean are skewed, Gouretski (2018) suggested to use adjusted Tukey's boxplot method for skewed distribution (Hubert & Vandervieren, 2008) to construct accepted local ranges of temperature and salinity for AutoQCs.

The demand in the scientific community for a unified and broadly accepted quality controlled (QCed) global archive has led to the establishing of the IQuOD – the international Quality-controlled Ocean Database initiative (Cowley et al., 2021). One of the primary IQuOD goals is to arrive on a consistent QC standard. The current work may be considered as a contribution to this joint effort and is focused on the following issues: 1) to present a new multiple-check AutoQC procedure which makes no assumption on the statistical distribution (we refer to this procedure as CODC-QC system (Chinese Academy of Sciences - Ocean Data Center (CODC) Quality Control system); 2) to conduct evaluation of the above procedure using two large manually controlled hydrographic datasets; 3) to apply the new procedure to the global hydrographic dataset and provide quality assessment for its distinct components; 4) to quantify the impact of the QC of temperature data on the ocean heat content (OHC) estimations.

As the shape of the temperature profiles is constrained by physical laws, checking the vertical structure of temperature profiles provides an additional possibility to identify spurious profiles. Existing AutoQC procedures imply a number of overall gradient checks which prove the shape of the temperature profile (e.g., (Garcia et al., 2018)). Here, we developed and evaluated an additional local vertical temperature gradient range check, where the local gradient ranges are constructed (similar to local temperature climatological ranges) without assuming the data distribution law.

Calculation of both temperature and temperature gradient climatological ranges at a certain geographical point assumes the ergodic hypothesis in which the average over time is assumed to be equal to the average over the statistical ensemble of data within some radius (bubble) around the point of interest. As another novelty of this study, the selection of data within the bubble takes into account water mass characteristics and topographic barriers. Application of the new QC scheme to two manually QC-ed benchmark datasets suggest the improved ability of the new scheme to identify spurious temperature profiles. It should be stressed here that the properly validated global hydrographic archive is crucially important for OHC estimation (Cheng et al., 2017), which is a primary indicator of human-induced climate change. We demonstrate that the choice of the QC scheme exhibits not a negligible impact on the estimation of

the long-term ocean warming.

In this study, we use a new hydrographic climatology complied in the Institute of Atmospheric Physics (IAP). Specifically, two by-products of this climatology are used: the local temperature climatological range (hereinafter IAP-T-range) and the local vertical temperature gradient climatological range (hereinafter IAP-TG-range). Both local climatological ranges are used to identify outliers in the observed temperature profiles.

This manuscript is structured as follows: the data used for this study are described in Section 2. The new CODC-QC procedure is outlined in the Section 3, with results being presented in section 4. In section 5, the performance of the CODC-QC is compared with the performance of several other QC procedures. In section 6, the application of the CODC-QC to the global WOD18 *in-situ* dataset is analyzed. The impact of QC on OHC estimation is illustrated in Section 7. Conclusions and discussions are presented in section 8.

2. Data

2.1 *in-situ* observations of temperature profiles

World Ocean Database 2018 (WOD18) (Garcia et al., 2018) is the main data source for this study. We used temperature profiles obtained between January 1940 and December 2020 by the following instrumentation types: Ocean Station Data (OSD), Conductivity-Temperature-Depth (CTD), eXpendable bathythermographs (XBT), mechanical bathythermographs (MBT), Argo profiling floats (PFL), drifting buoys (DRB), autonomous pinniped bathythermographs (APB), mooring buoys (MRB), undulating oceanographic recorder (UOR), Glider (GLD). In total, 16,804,361 temperature profiles were selected for the analysis. The data coverage in the North Polar region was improved by incorporating additional 52,253 non-WOD CTD profiles obtained from the Alfred-Wegener-Institute, Bremerhaven, Germany, the Department of Fisheries and Oceans of Canada, the Freshwater Institute, Bedford Institute of Oceanography, Institute Maurice-Lamontagne, Northwest Atlantic Fisheries Centre, and the Institute of Ocean Sciences.

2.2 Benchmark profile datasets

Two temperature profile datasets which undergone the ExpertQC were used as benchmark datasets for this study: the QuOTA dataset and the WOCE one-time dataset.

(1) QuOTA dataset

The Quality-controlled Ocean Temperature Archive (QuOTA) obtained from the Commonwealth Scientific and Industrial Research Organization (CSIRO, Australia) (Gronell & Wijffels, 2008; Wijffels et al., 2008) is used as a benchmark dataset for QC procedure performance evaluation. The QuOTA dataset is a collection of temperature profiles from the Indian Ocean obtained between 1772-2005. The dataset is used as a benchmark because nearly 30% of the data were

entirely under the expert (manual) QC. Among the total of 14,424,324 temperature measurements (36,785 profiles), 83.39% were flagged as good (flag=0,1,2,5), and the remaining are flagged as bad (flag=3,4). About 80% of profiles in this dataset are recorded by XBT, with the rest being mainly MBT and bottle profiles. Respectively, the overall quality flag statistics characterize rather the XBT profiles.

(2) WOCE one-time dataset

The World Ocean Circulation Experiment (WOCE) conducted under the auspice of the World Climate Research Program (WCRP, 1988a, 1988b) provided the global-scale hydrographic dataset of outstanding quality, because individual efforts were controlled and orchestrated by the WOCE hydrographic office, guaranteeing the high and uniform accuracy of the data (Gouretski & Jancke., 2001; Gouretski & Koltermann, 2004). The CTD profiles from WOCE One-Time Survey (King et al., 2001) were used for our purpose. In total, 8,793 profiles (18,294,163 temperature measurements) between 1985-1997 were collected, with nearly all data (18,052,052 measurements, 98.68%) are flagged as good (flag=2). Due to the extremely high data quality this dataset is especially useful for the evaluation of how well a QC procedure works in retaining good data, rather than detecting the outliers.

3. CODC-QC automated quality control procedure

3.1 CODC-QC quality control checks

Similar to other QC schemes, our CODC-QC consists of a series of quality checks (Table. 1). Among the total of 13 checks, Checks 1-4 are not related to any variables measured at depths. Checks 5-7 are related to the acceptable range of the measured variable (temperature). Checks 8-12 are related to the vertical structure (shape) of the profile. Checks 13 are applied only to profiles obtained by specific instruments. In the following, the distinct checks are described in detail.

In the CODC-QC, the definition of the QC flag is dichotomic for each check at each observed depth, with 0 denoting the acceptable (good) value and 1 denoting the rejected (bad) value. Additionally, a final quality flag is provided based on all distinct quality checks. The user can use the final quality flag or make personal decision on the quality flag based on all distinct quality checks. For more information about the CODC-QC system, please refer to the online User Manual Document.

Table 1. Details of each Quality Control (QC) check of CODC-QC.

Order	Name of checks	Comments
(1)	Basic information check	Check whether date, time, location are in acceptable ranges.
(2)	Sample level order check	Check whether profile sample depths increase monotonically
(3)	Instrument maximum depth check	Check weather sample depth exceeds instrument type maximum
(4)	Local bottom depth check	Check whether sample depth exceeds local bottom depth

(5)	Profile range check	Global range check
(6)		Sea-water freezing point check
(7)		Local climatological range check
(8)	Profile shape check	Constant value check
(9)		Spike check
(10)		Multiple extrema check
(11)		Global vertical gradient check
(12)		Local gradient climatological range check
(13)	Instrument specific check	Only checks in XBT profiles related to wire stretch, leakage, hit

(1) Basic information check

In case the year and/or date or geographical coordinates are missing or exceed permitted range all levels of such profiles are flagged. This check has been already applied in several existing QC schemes (e.g., Catherine et al. (2014); Good et al. (2013); Intergovernmental Oceanographic Commission (2010); U.S. Integrated Ocean Observing System (2015)).

(2) Sample level order check

It is common in the oceanographic practice to have monotonically increasing sample depth of a hydrographic profile. The check proves whether sample depth monotonically increases and flags all observations in case it does not.

(3) Instrument maximum depth check

Each instrument type is designed to operate within a certain depth range. For instance, MBT cannot withhold pressure below 320 meters, and the maximum sample depth for XBT probes is limited by the length of the wire which depends on the probe type (Boyer et al., 2019). If an observed depth value falls outside the nominal maximum depth for the corresponding instrumentation type, the values below the nominal maximum depth are flagged. Additionally, for the CTD and XTB profiles the depth levels above 2 and 3 meter respectively are flagged according to the indication by Good et al. (2013).

(4) Local bottom depth check

This check identifies whether the deepest sampled level is larger than the local bottom depth. The latest version of global 0.5 arc-second resolution digital General Bathymetric Chart of the Oceans (GEBCO) (Tozer et al., 2019) provides the estimate of the local ocean bottom depth (D). Because of errors in the digital bathymetry (see Fig. S3 in the supporting information for more details) and due to the uncertainty in the profile location, a certain tolerance (d) is permitted: $d = 30.0 + D * 0.087$ for D $<$ 1000m; $d = 80.0$ for D $>$ 1000m for the profile located north of 60°S. Since the digital bathymetry is less reliable around Antarctica, $d = 270 - D * 0.37$ south of 60°S. Measurements at levels deeper than $D+d$ are flagged. Cases where the observed depth exceeds the bottom depth may occur due to the wrong geographical position assignment (more

typical for old historical profiles) or because some instruments (XBT, MBT) continue data transmission after hitting the bottom.

(5) Global range check

The check aims to identify observations where temperature is grossly in error. The global depth-dependent temperature range is based on all available profiles. It is constructed using the upper and the lower boundaries of the 0.5% and 99.5% quantiles respectively. Similar range check was implemented in Intergovernmental Oceanographic Commission (2010). Values exceeding the overall depth-dependent range are flagged. For several regions of the world ocean having specific thermohaline structure the overall global ranges were adjusted (e. g. the Mediterranean Sea, Black Sea, Sulu Sea).

(6) Sea-water freezing point check

In addition to the global range check, observed temperature values are compared with the sea water freezing point temperature defined by the following formula (Intergovernmental Oceanographic Commission, 2010):

$$T_f = -0.0575 * S + 0.001710523 * S^{\frac{3}{2}} - 0.0002154996 * S^2 - 0.000753 * P \quad (1)$$

Where T_f is the freezing point temperature in degrees Celsius, S is the salinity in PSU and P is the pressure in decibars. If the accompanying observed salinity profile is absent (e.g., for XBT, MBT, APB, GLD temperature profiles), the monthly climatological salinity reference from Cheng et al. (2020) was used. Levels with temperatures smaller than T_f are flagged.

(7) Local climatological range check

Whereas the global range check identifies observations which are grossly in error, e.g., the so-called blunders, the local climatological range check proves whether the observed values fall within the acceptable local monthly temperature climatological range (i.e., IAP-T-range). The local ranges are constructed for the nodes of the 1-degree box. Similar to the global temperature range, the 0.5% and 99.5% quantiles define the minimum and the maximum local temperature values at each depth, with details given below in the Section 4. All observations that fall outside the local monthly depth-dependent temperature range are flagged.

(8) Constant value check

Some profiles are characterized by unrealistically thick layers where temperature does not change with depth. In case of electronic instruments such apparent thermostads can be caused by the malfunctioning of the temperature sensor (CTD) or reversing thermometers (old Nansen casts) leading to "stuck values" on the profile. All measurements within the detected thermostads are flagged.

(9) Spike check

Spikes (e.g., temperature values strongly deviating from the neighbor measurements above and below) occur typically due to the malfunction of electronic instruments like CTD or XBT. In case of mechanical instruments (Nansen casts, MBT) spikes occur mostly due to operators' errors. Spike check has been implemented in several QC schemes (Cabanes et al., 2021; Intergovernmental Oceanographic Commission, 2010; Liu et al., 2017; Wong et al., 2020). In the CODC-QC, for the three adjacent temperature measurements (denoted as T_k, T_{k+1}, T_{k+2}), we calculate:

$$S = \{T_{k+1} - (T_k + T_{k+2}) * 0.5\} - (|T_{k+2} - T_k| * 0.5) \quad (2)$$

If S is greater than the depth-dependent thresholds, T_{k+1} is flagged and defined as a spike. In addition, this spike check is not performed for the profiles with low vertical resolution (see Gouretski (2018) for details).

(10) Multiple extrema check

This check identifies profiles with unrealistically big number of local temperature extrema being similar to that used by Gouretski (2018). The selection of the threshold (e.g., of the local maximum/minimum range) considers the measurement precision and the typical scattering due to the fine temperature micro-structure. All levels of profiles failing the check are flagged.

(11) Global gradient range check

This check identifies pairs of levels for which the vertical temperature gradient exceeds the overall depth-dependent range, which is similar to the global range check for temperature. Because vertical distribution of temperature with depth is not-linear, the finite-difference estimates of the vertical gradient depends on profile vertical resolution, with the magnitude of the calculated gradient decreasing with the increase of the vertical gap between the two layers. Therefore, unlike other schemes (e.g., Garcia et al. (2018)), we apply different gradient ranges depending on the profile vertical resolution. We calculate the temperature gradient (∇T) at level k using central differences as follows:

$$(\nabla T)_k = \frac{\mathrm{T}}{\mathrm{Z}} = \frac{\frac{1}{n_2}\sum_{i=k}^{n_2} T_i - \frac{1}{n_1}\sum_{i=k-n_1}^{k} T_i}{Z_{k+n_2} - Z_{k-n_1}} \quad (3)$$

Some instrumentation types provide temperature profiles which resolve the fine thermal structure of the water column characterized by the high values of the temperature gradient. Here, we are only interested in detecting the gradient outliers for the vertical structures of a larger scale (i.e., scales larger than characteristic scales of the ocean fine thermohaline structure). This is achieved by choosing the smallest values of n_1 and n_2 in Equation (3) which corresponds to the vertical gap (Z) greater or equal to 10 meters empirically. The temperature at level k for which the vertical gradient falls outside the global depth-dependent gradient range is flagged.

(12) Local gradient climatological range check

This check is identical to the global gradient range check but with gradient

ranges defined locally similar to the local climatological range check for temperature described in (7). Here, the IAP-TG-range is used as the local climatological vertical gradient range (see Section 4 for details).

(13) Instrument specific checks

In the highly non-homogeneous archive like WOD, spurious temperature profiles often exhibit features which are characteristic only to specific instrumentation types. These features can be attributed to specific problems encountered during the data acquisition. In this check, quality issues of XBT temperature profiles related to wire stretch, wire insulation damage, leakage, wire noise or bottom hit are detected. As suggested by Barton and Gonzalez (2016), if a temperature measurement at a specific level was linked to one of these problems, the measurements below this level should be all flagged as bad.

3.2 Evaluation of the CODC-QC system

Currently, there are no unified standards to evaluate the performance of different QC schemes. The International Quality Controlled Ocean Database (IQuOD) and Argo science teams provided useful practices (Cowley et al., 2021; Domingues & Palmer, 2015; Roemmich et al., 2019; Wong et al., 2020). In this study, we used two manually QC-ed datasets to evaluate the performance of the new CODC-QC system of this study.

The dichotomous metrics of True Positive Rate (*TPR*), False Positive Rate (*FPR*) and True Negative Rate (*TNR*) first proposed by Yerushalmy (1947) are used for the QC performance evaluation. For each measurement, the benchmark dataset provides truth pass and truth reject rates, and the AutoQC results (pass and reject) are compared with the values for the benchmark dataset. *TPR* characterizes how well the AutoQC detects bad data, while *TNR* and *FPR* characterize the ability of the scheme to retain good data. The best AutoQC procedure performance is achieved when the *TPR* and *TNR* are as high as possible, and the *FPR* is as low as possible. The metrics are applied to the observations at each observed level, rather than to entire observed profiles.

Additionally, the Open-sources Python project developed by IQuOD (https://github.com/IQuOD/AutoQC) was used to compare the performance of the CODC-QC with AutoQC procedures from four other institutions (NOAA's Atlantic Oceanographic and Meteorological Laboratory QC procedure (AOML-QC; version 2018), Argo real-time QC procedure (Argo RTQC; version 2.5), Commonwealth Scientific and Industrial Research Organization QC procedure (CSIRO-QC), and Integrated Climate Data Center of the University of Hamburg QC procedure (ICDC-QC)) with using the above dichotomous metrics.

4. Local climatological ranges for temperature (IAP-T-range) and for the vertical temperature gradient (IAP-TG-range)

4.1 Definition of the acceptable range

It will be shown below that checking the observed data against the local climatological ranges for temperature and vertical temperature gradient (further call for brevity temperature gradient) is most effective in identifying outliers compared to other quality checks. As discussed in Section 1, acceptable ranges for observed variables are often constructed under the assumption of Gaussian distribution with Tukey's box-plot method (McGill et al. 1978), which is often being applied to define the acceptable range. However, local distributions of temperature and temperature gradient are typically skewed. The temperature skewness maps for selected levels (Fig. 1a, c) show the largest skewness magnitude in the boundary currents regions (Fig. 1g), within the Antarctic Circumpolar Current (ACC) belt (Fig. 1e) and within the tropical and equatorial zones. The temperature gradient skewness maps (Fig. 1b, d) are patchier and demonstrate different patterns, with prevailing negative values (Fig. 1h) except for the polar regions where temperature inversions occur (Fig. 1f).

Assumption of the symmetry of statistical distribution results in the undesirable property of a QC to flag too many good data as outliers. Gouretski (2018) used the adjusted boxplot method modified for skewed distributions (Vanderviere & Huber, 2004). The QC scheme inter-comparisons study by Good et al (submitted) demonstrated significant improvement in the ability of that scheme to identify outliers compared to the schemes assuming Gaussian distribution (e.g., WOD-QC and AOML-QC procedures). However, as noted by Adil and Irshad (2015), the adjusted boxplot method works well only for highly skewed distributions ($|\text{skewness}| \geq 3$), but produces fences larger than the extremes of the data for smaller skewness values.

Here we present a different approach to construct the acceptable range without any assumption on the statistical distribution law and define the lower fence as the upper boundary of the 0.5% quantile and the upper fence as the lower boundary of the 99.5% quantile. These boundaries represent the local climatological temperature range (IAP-T-range) and the local climatological vertical temperature gradient range (IAP-TG-range).

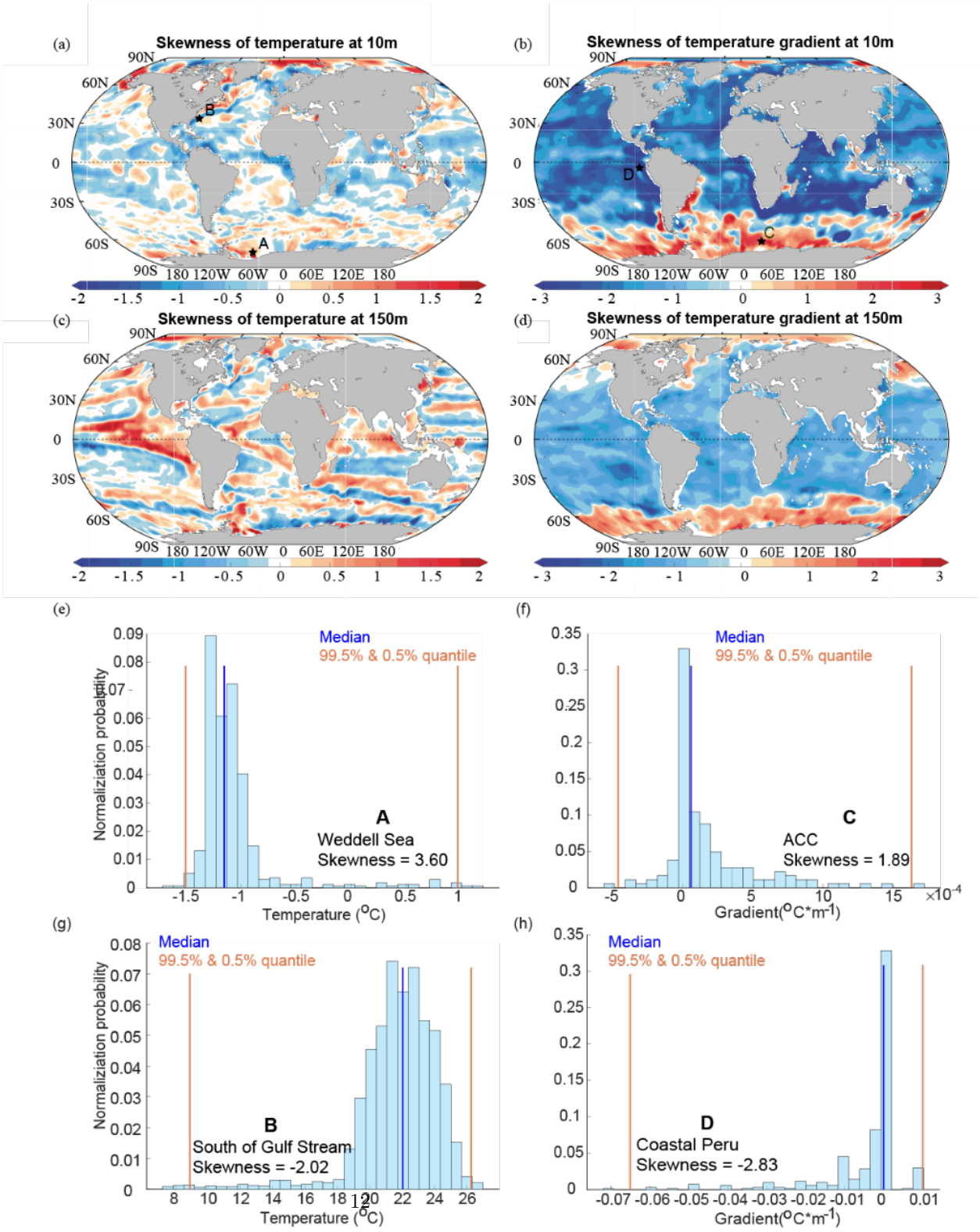
(a)
Skewness of temperature at 10m
(b)
Skewness of temperature gradient at 10m
(c)
Skewness of temperature at 150m
(d)
Skewness of temperature gradient at 150m
90N
60N
30N
0
30S
60S
90S
180 120W 60W 0 60E 120E 180
A
B
C
D
(e)
(f)
(g)
(h)
Median
99.5% & 0.5% quantile
Normaliziation probability
Temperature (°C)
Gradient(°C*m-1)
A
Weddell Sea
Skewness = 3.60
B
South of Gulf Stream
Skewness = -2.02
C
ACC
Skewness = 1.89
D
Coastal Peru
Skewness = -2.83

Fig. 1 (a, c) The skewness of the temperature distribution in one-degree boxes at 10m depth (January); (b, d) same as (a, c), but for the vertical temperature gradient (July). The maps are based on all quality-controlled WOD18 temperature profiles between 1940 and 2020; (e-h) temperature histograms for four selected locations: (A) the Weddell Sea, (B) south of Gulf Stream, (C) the Antarctic Circumpolar Current (ACC), (D) off the Peru coast. The locations are marked as black starts on maps in (a) and (b).

4.2 Initial data QC and interpolation

Temperature data discussed in Section 2.1 were used to construct local climatological ranges for temperature (IAP-T-range) and temperature gradient (IAP-TG-range). First, some crude QC checks was performed excluding observations fail the basic information check, the instrument depth check, the local bottom depth check, the global range check, levels order check, instrument maximum depth check, global gradient range check, spike check, and constant value check (see Section 3). Secondly, we interpolated temperature profiles to 119 standard levels from surface to 6000m (1 m, 5–100 m in 5m intervals, 110–200 m in 10m intervals, 220–400 m in 20m intervals, 425–700 m in 25m intervals, 750–2000 m in 50m intervals, and 2100–6000 m in 100m intervals) using the parabolic interpolation (Reiniger and Ross 1968). The interpolation was not performed where the interval between two adjacent levels exceeds a certain depth-dependent threshold followed by Gouretski (2018).

4.3 Water mass dependent data selection within the local influence bubble

The local climatological ranges for temperature and temperature gradient are calculated on 1-degree boxes at each standard level. For each grid point, the surrounding profiles are selected within the 555 km radius (bubble) to guarantee the sufficient number of profiles even in the data-gap regions (same as in Gouretski (2018)), and 555 km is consistent with spatial decorrelation length scale for inter-annual scale variability. The data selection is performed on the monthly bases, with the required minimum number of profiles in the local bubble set as 40 in the upper 250m, 30 between 250 and 450m, and 20 between 450 and 2000m (these thresholds are chosen empirically). If the number of profiles for a climatological month is less than the above thresholds, the profiles from the adjacent months are also used.

Except for some spatially homogeneous areas of the global ocean, selecting all data within a bubble lead to the undesirable increase of the variable ranges because it might include profiles from different water masses. Here, we developed an additional selection algorithm which takes into account the anisotropy of the temperature distribution within the influence bubble. Firstly, we calculate temperature and temperature gradient monthly means and standard deviations ( ) by all temperature profiles in each 1-degree box (background values in Fig.2a). Secondly, for the central box of each influence bubble, we define the local temperature range as [mean±] (Fig.2b-d). Because of the higher variability of the

temperature gradient, the gradient range for the center of the bubble is defined as [mean±0.3*] (0.3 is an empirical choice). Thirdly, the boxes within the influence bubble are ordered according to their mean values and the boxes whose mean values fall within the range of the central box are retained (Fig.2a-d) for the further analysis.

The examples of the data selection are shown in Fig. 2 for three locations in the Pacific Ocean east of Japan. The box-mean temperatures in ascending order are shown for each bubble along with the respective standard deviation (Fig.2b-d). The histograms (Fig.2e-g) characterize the anisotropy of temperature distribution within each bubble. The northernmost location (Bubble A) is within the cold Oyashio current, the southernmost location is within the Kuroshio extension (Bubble B), the central location (Bubble C) is within the Subarctic front. In all three cases the water-mass dependent profile selection excludes the northernmost and the southernmost 1°×1°-degree boxes within the bubbles, retaining only the boxes with water mass characteristics similar to those of the central box. For instance, water temperature at the central point of Bubble A is dominated by the cold Oyashio current flowing from the north-east, so that predominantly the boxes from the northern half of the bubble are selected. At the center of the Bubble B the Kuroshio Extension waters flowing in the zonal direction dominate, so that almost all boxes in west-east direction are retained except for the southernmost and the northernmost parts of the bubble. Figs.2e-g show that the statistical distributions in bubbles A and B are highly skewed. These examples illustrate that the additional selection of data is powerful to better estimate the local climatological ranges because of the water mass properties are taken into account.

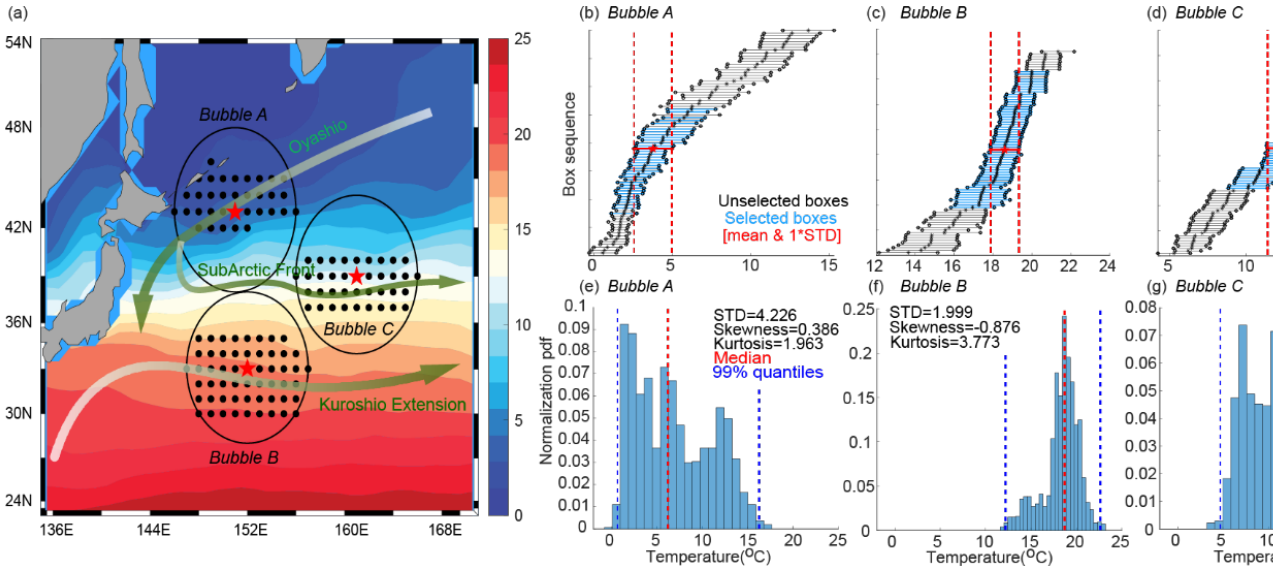

Fig. 2 Illustration of the water mass dependent data selection method for

three locations east of Japan: a) climatological temperature map at 10m depth level in January and bubbles (A-C) with centers of the selected 1x1-degree boxes (black), location of the central box is shown in red; (b-d) box mean temperatures with respective within box standard deviations in ascending order for three bubbles, vertical red lines denote the temperature range for box-selection; (e-d) histograms of the temperature distribution within each bubble, vertical blue lines indicate the local temperature climatological range for the new CODC-QC system developed in this study.

4.4 Accounting for topographic barriers

As discussed in Shahzadi et al. (2021), artefacts due to neglecting of spatial discontinuities are observed in some ocean climatologies. Fig.3 shows two examples of topographic barriers. Fig. 3a-c shows that although the data from the selected boxes are statistically identified as the same water mass using the procedure described in Section 4.2-4.3, these boxes are located within two currents: the Aleutian North Slope Current and the Alaska Coastal Current, with the Alaska Peninsula as a topographic barrier between them. For the bubble immediately south of Central America (Fig.3d-f), the statistical procedure chooses the data both in the Atlantic and Pacific oceans which are disconnected from each other. To cope with such situations, we take into account the distribution of the bottom barriers between the bubble central and selected profiles and retain only those profiles for which no topographic obstacle between the central bubble point exists.

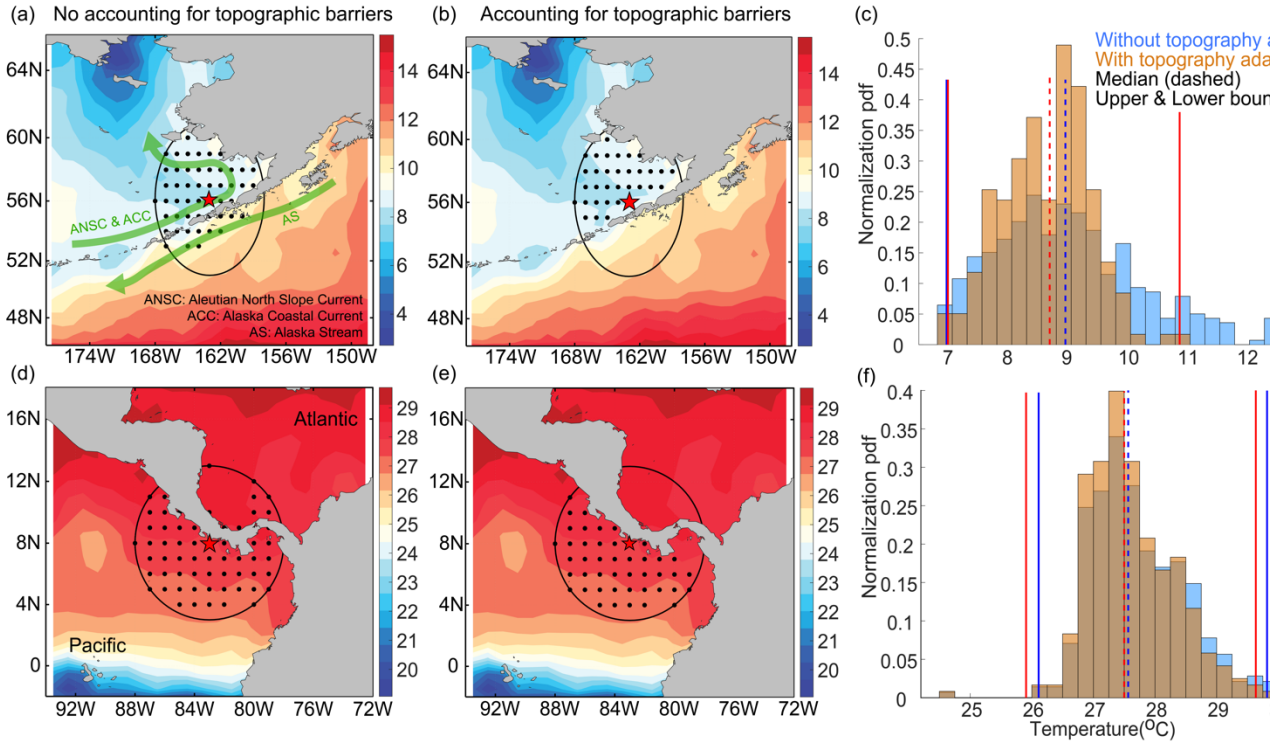

Fig. 3. Two examples of box-selection without (a, d) and with (b, e) the account for topographic barriers: data selection bubble north of Alaska Peninsula (a-b) and south of Central America (d-e). Monthly climatological temperature distribution at 5m depth (January) is shown in the background; (c, f) shows temperature histograms for data selection with and without account for topographic barriers.

4.5 Global fields of the local climatological range for temperature (IAP-T-range)

Using vertically interpolated temperature profiles (see Section 4.2), the following set of parameters was calculated for temperature: mean, median, standard deviation, 99.5% quantile, 0.5% quantile, skewness, and kurtosis. Altogether these parameters represent the IAP-T-range climatology, with quantile boundaries corresponding to the local climatological ranges. Calculations were performed on the 1°×1° grid at 119 levels. For each 1°×1° field, a nine-point mean filter was applied to further exclude some outliners and ensure spatial coherency. For temperature monthly climatological ranges are available for the layer 0-2000m and seasonal ranges for 2000-4000m. Below 4000m, ranges are calculated using all available data.

Fig. 4a-h shows the fields of IAP-T-range with local minimum and maximum climatological temperature for four selected representative depth levels (The

local temperature climatological mean, median, standard deviation fields are shown in Fig. S1 in the supporting information). The respective ranges (T_{max}-T_{min}) are shown in Fig.4i-l. At these depth levels, the maximum range is found in the high-energetic regions of the ocean corresponding to the western boundary currents like Gulf Stream, Kuroshio, Antarctic Circumpolar Current and to the equatorial belt in the Pacific Ocean. Broad temperature range at the level 1600 m in the North Atlantic Ocean is linked to the spreading of the Mediterranean water. The large-scale patterns are in agreement with previous studies (e.g., Antonov et al. (2004); Schmidtko et al. (2013) and Gouretski (2019)).

The performance of the local temperature climatology range check, driven by the IAP-T-range, is illustrated by means of profiles from six locations in the World Ocean (Fig. 5). Profiles were obtained by means of different instruments. Among the six examples in two cases (Fig. 5a, c) all observations fall outside the local climatological limit. In other cases, only a part of measurements is flagged as outliers. For instance, the MBT profile (Fig. 5b) is characterized by a shift below 50m level, with all data flagged below 60 m. Similar situation can be seen for the XBT profile (Fig. 5f), where shift at ~75 m depth might indicate wire stretching. For two Argo profiles (Fig. d,e), the erroneous data were reported only at a number of levels.

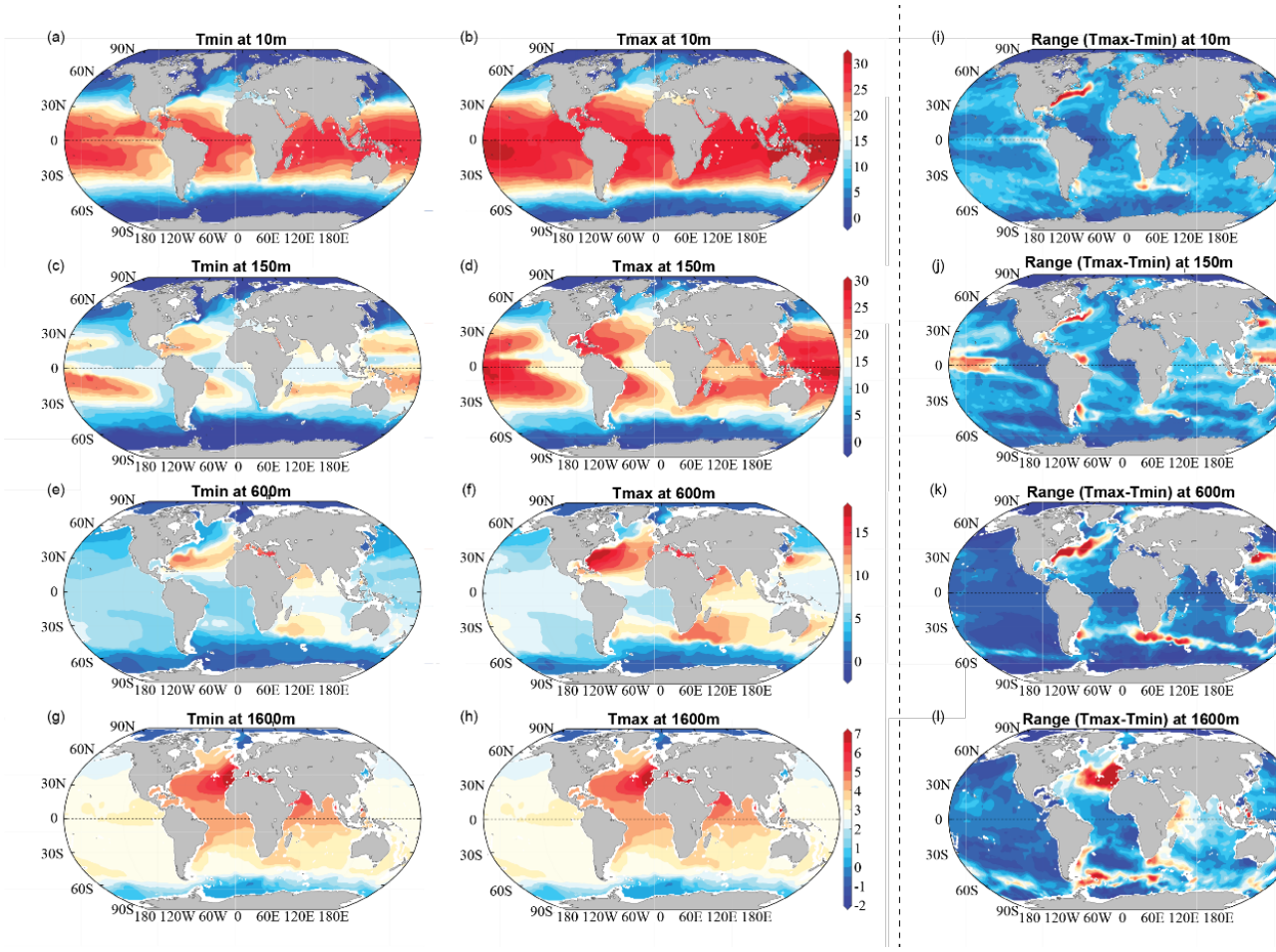

Fig. 4. The IAP-T-range temperature climatology (January) fields at several representative levels including 10m, 150m, 600m, 1600m depths: (a, c, e, g) local minimum values (T_{min}), (b, d, f, h) local maximum values (T_{max}), and (i-l) local temperature climatological range (T_{max}-T_{min}).

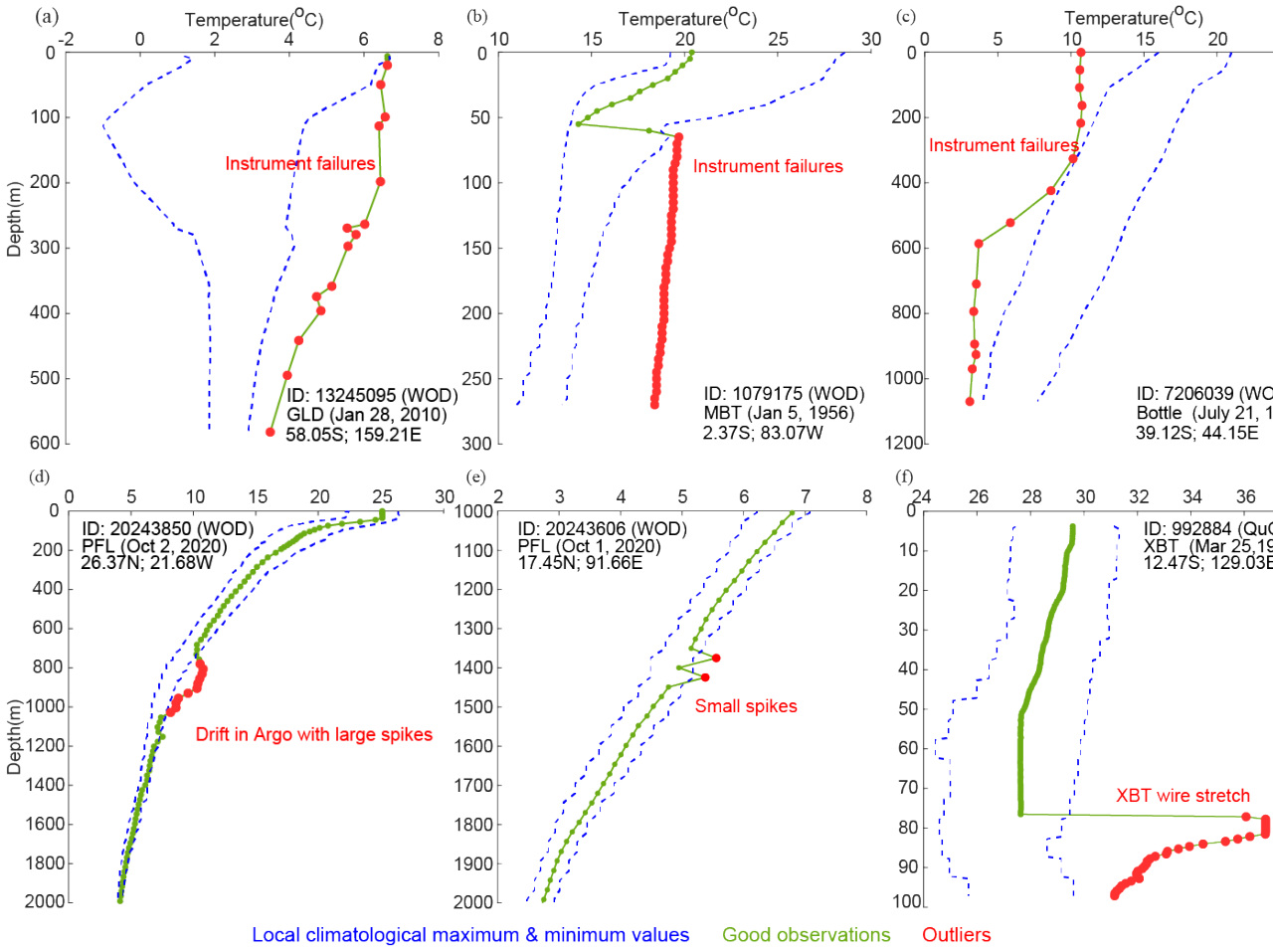

Fig. 5. Performance of the local climatological range check as applied to a number of temperature profiles. The local climatological maximum and minimum values (IAP-T-range) are shown with blue dashed lines. Temperature profiles are shown in green (good values) and red (bad values).

4.6 Global fields of the local climatological range for the vertical temperature gradient (IAP-TG-range)

Similar to the global fields of IAP-T-range discussed in Section 4.5, the 99.5% and 0.5% quantiles of temperature gradient within the local influence bubble are set as the upper and lower climatological limits. In a deviation from temperature ranges for the vertical temperature gradient were calculated only within the layer 0-2000m. Since temperature distribution with depth is not linear, the estimation of the vertical temperature gradient depends on the vertical spacing between the two levels where the gradient is calculated. Increasing vertical spacing between the observed levels leads in general to the decrease of the gradient magnitude

(because the noise will be exacerbated when calculating derivatives using higher resolution profiles). Respectively the gradient climatological ranges used for the QC of profiles must take into account the spacing between the observed levels. We note that the vertical resolution (Z) differs considerably among different instrumentation types. For instance, the common vertical resolution for CTD profiles in the WOD archive is 2 decibars, whereas the historical Nansen bottle casts rarely have observations at more than 15-20m, with spacing between the observed levels typically increasing with depth from 5-10m near the surface to several hundred meters in the lower part of the profile. Fig. 6 shows the overall IAP-TG-range versus depth for five values of the vertical spacing.

Local minimum and maximum temperature gradient climatological values, along with the local range are shown in Fig.7 for four selected levels (The local temperature gradient climatological mean, median, standard deviation fields are shown in Fig. S2 in the supporting information). For the uppermost level (10 m), a clear asymmetry between the Northern and the Southern Hemispheres can be seen both in minimum and maximum values, with much higher gradient magnitudes in the Northern hemisphere. At a deeper level (160m), the asymmetry in two hemispheres vanishes, with the maximum width of the range occurring in the near-equatorial belt. This is mainly because of the water column change induced by the switch between El-Nino and La-Nina regimes. The pattern of the gradient range becomes patchier with increasing depth. At 600m level, the widest range is found mostly off the eastern coasts of the Atlantic, Indian and Pacific oceans. There is a broad agreement of the IAP-TG-range climatology with a number of existing studies (Luyten et al., 1983; Neal et al., 1969; Oakey & Elliott, 1977; Wang et al., 2000).

Similar to Fig. 5 for temperature, Fig. 8 provides several examples of the temperature gradient check performance. The gradient range check identifies some spurious temperature inversions (Fig.8b) and spikes (Fig. 8 c,d) that could not be detected by the local temperature climatological range check or global vertical gradient check.

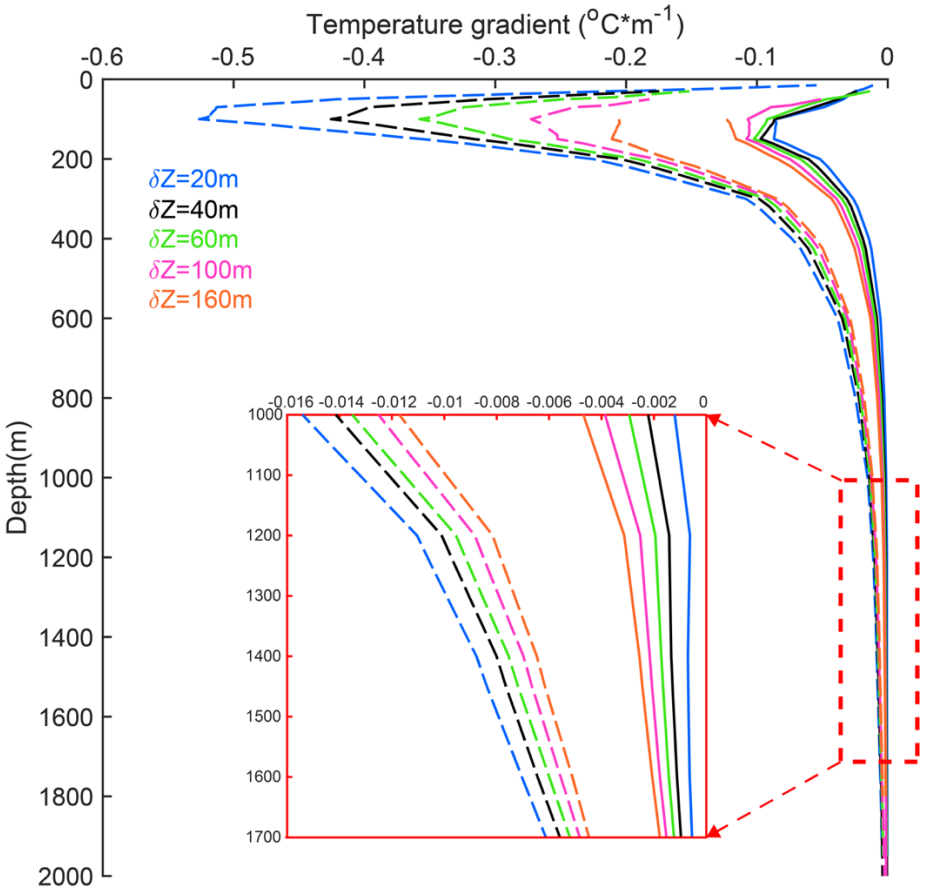

Fig. 6. Global vertical temperature gradient range (IAP-TG-range) for July as a function of vertical spacing (Z) indicated by different colors. The dashed lines indicate the lower bound and the solid lines indicate the upper bound.

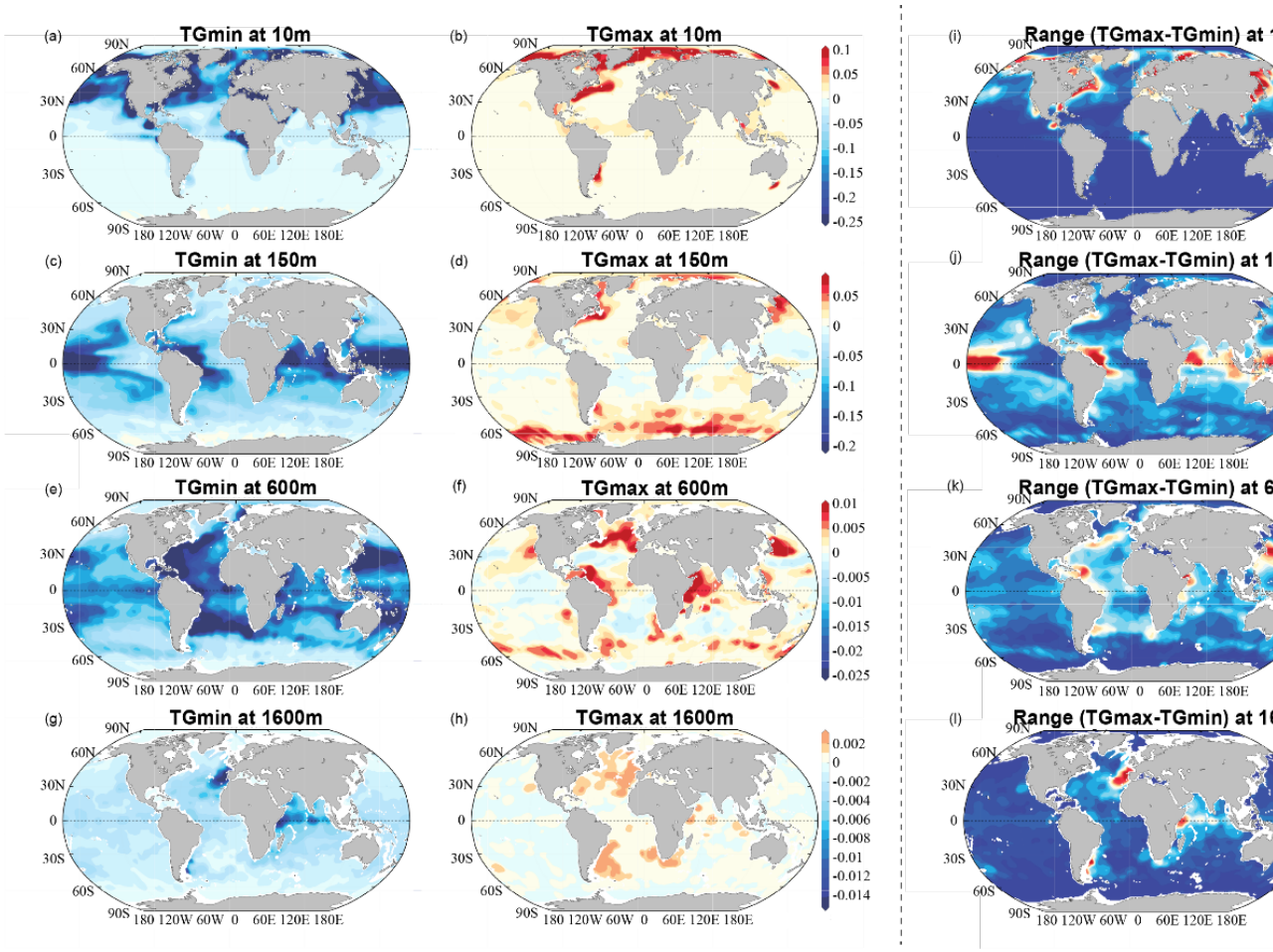

Fig. 7. The IAP-TG-range temperature gradient climatology (July) at 10m, 150m, 600m, 1600m depth levels: (a, c, e, g) the local minimum value (TG_{min}); (b, d, f, h) the local maximum value (TG_{max}), and (i-l) the local temperature gradient climatological range (TG_{max}-TG_{min}). All fields correspond to the vertical spacing Z=10m.

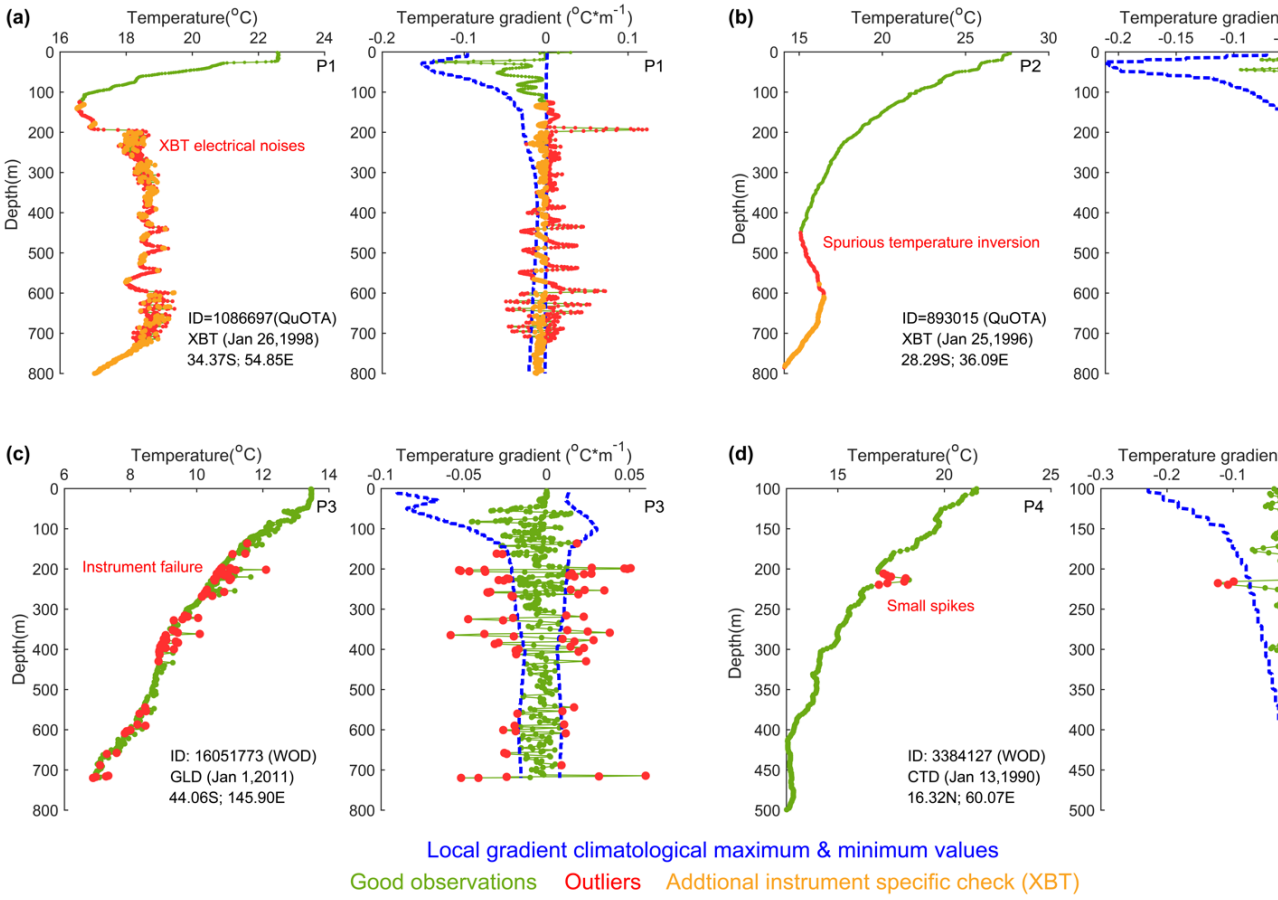

Fig. 8. Performance of the local temperature gradient climatological range check as applied to four temperature profiles. The local minimum and maximum gradient values (IAP-TG-range) are shown in blue. Temperature or the corresponding vertical temperature gradient are shown in green (good values) and red (bad values).

5. The evaluation of the CODC-QC procedure performance using benchmark datasets

We used two benchmark datasets (QuOTA and WOCE), which have been rigorously QC-ed by experts (see sections 3.1, 3.4 for details) to evaluate the performance of the CODC-QC. Table 2 summarizes the *TPR*, *FPR* and *TNR* for the new CODC-QC and other four QC systems, whereas Table 3 summarizes the contribution of different CODC-QC quality checks to the benchmark metrics. The CODC-QC demonstrates the highest *TPR* (~67%) in the QuOTA dataset, mainly attributed to the local climatological range check and local gradient climatological range check. Note that QC system from Atlantic Oceanographic and Meteorological Laboratory (AOML-QC) applies climatological mean plus/minus

4-sigma range from World Ocean Atlas 2018 (WOA18) (Locarnini et al., 2019) to define the local range, which yields the second highest TPR of 40.22%. For the WOCE dataset, the *TNR* of the CODC-QC is ~98%, meaning that only 2% of good data are flagged as bad by the QC procedure. This 2% false rejection rate (~5% for the QuOTA dataset) is mainly attributed to the local climatological range checks (see Table 3).

According to Table 3, the local temperature climatological range check is the most effective part of the QC procedure (highest TPR equal to 51.00%). The local temperature gradient climatological range check and the instrument specific check play a secondary role. However, these three checks also contribute the largest FPR (~1% in total). In total, the CODC-QC can detect 66.97% bad data in the QuOTA dataset, with a trade-off of rejecting 5.45% good data in the QuOTA dataset and retaining 98.09% good data in the WOCE dataset. These results suggest that the local climatological range check and the local vertical gradient range check are the most important modules in the CODC-QC, consistent with the previous understanding for example in Gouretski (2018).

It should be noted that both WOCE and QuOTA benchmark datasets have their limitations. WOCE dataset is represented (in our case) only by the CTD profiles, and QuOTA dataset includes only XBT, MBT, and Bottle data. Both datasets represent only a minor fraction of the entire global hydrographic archive and do not cover the Arctic Ocean and many marginal or closed oceanic basins.

Table 2. The *TPR*, *FPR* rates for the QuOTA dataset and *TNR* rate for the WOCE dataset for five QC systems.

QC institutes	QuOTA	WOCE	
	TPR	FPR	TNR
CSIRO-QC	16.28%	0.32%	99.99%
ICDC-QC	33.17%	0.89%	90.77%
Argo RTQC	2.44%	0.06%	99.99%
AOML-QC	40.22%	2.17%	45.73%
CODC-QC	66.97%	5.45%	98.09%

Table 3. The *TPR, FPR* and *TNR* rates for distinct quality checks of the CODC-QC for the QuOTA and WOCE benchmark datasets.

QC checks of CODC-QC	QuOTA	WOCE	
	TPR	FPR	TNR
Basic information check	/	0%	100%
Sample level order check	/	0%	100%
Instrument maximum depth check	0.01%	0.01%	99.94%
Local bottom depth check	1.21%	0.06%	99.93%
Global range check	17.70%	0.02%	99.99%

QC checks of CODC-QC	QuOTA	WOCE	
Freezing point check	0.06%	0.01%	99.99%
Local climatological range check	51.00%	1.88%	98.96%
Constant value check	4.07%	0.01%	99.99%
Spike check	0.11%	0%	100%
Multiple extrema check	9.55%	0.71%	100%
Global vertical gradient check	21.34%	0.04%	99.99%
Local gradient climatological range check	44.03%	2.95%	99.49%
Instrument specific check (XBT)	43.82%	2.88%	100%
TOTAL	66.97%	5.45%	98.09%

To further illustrate the capability of the CODC-QC in detecting bad data, we provide composite plots of temperature profiles before and after the application of the CODC-QC and ICDC-QC procedures to the same arbitrarily selected profiles from the QuOTA dataset (Fig. 9). The raw profiles (Fig. 9a) include many erroneous observations distributed over the whole depth-temperature space. Comparison shows that the CODC-QC is superior in the ability to detect outliers compared to the ICDC-QC (Fig. 9 c,d), with the subset of the retained good observations being very similar to the benchmark data after the ExpertQC (Fig. 9b)

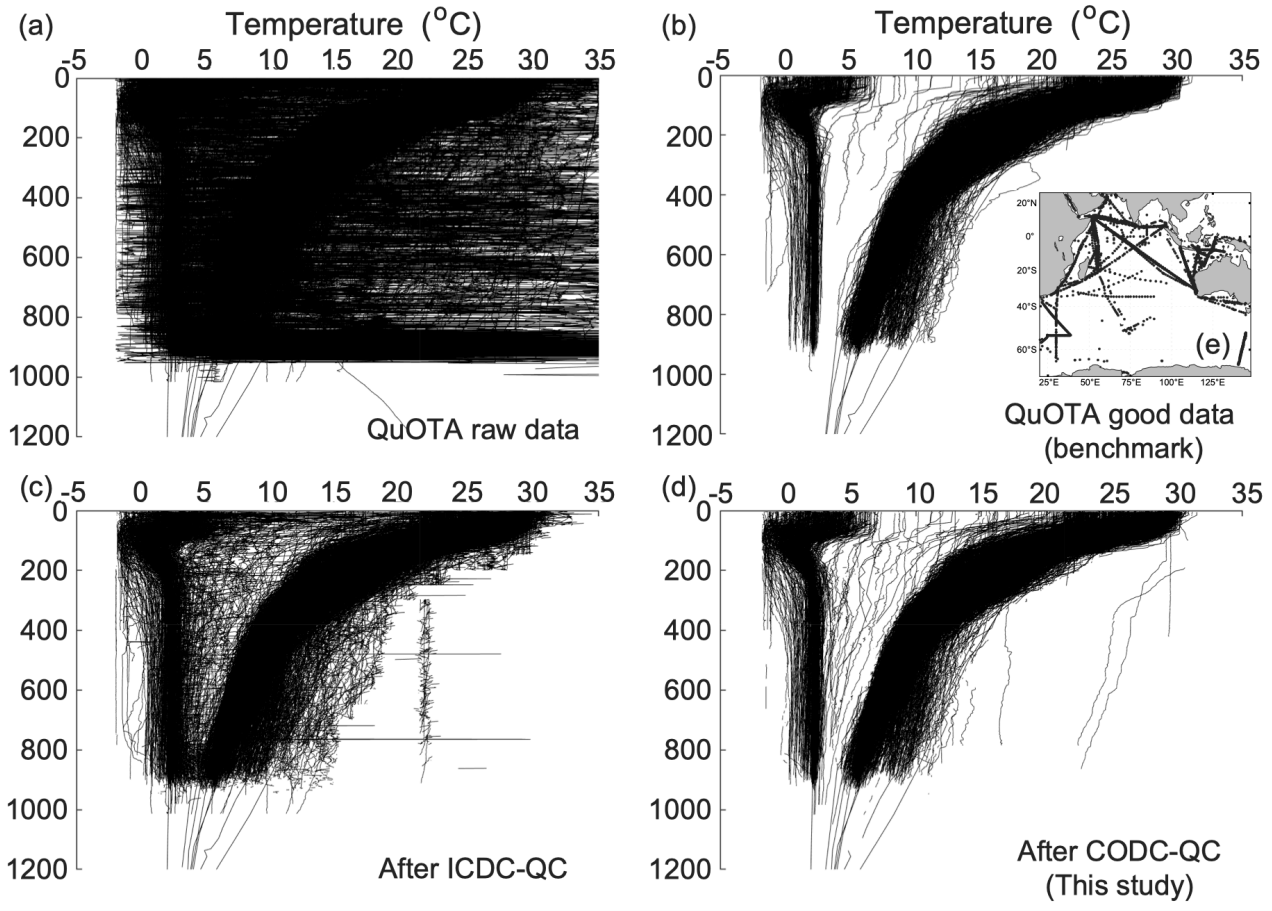

Fig. 9. Application of three AutoQC procedures to 3000 temperature profiles arbitrarily selected from the QuOTA dataset: (a) raw profiles; (b) profiles after removing data flagged by manual/expert QC; (c) same as (b) but for the ICDC-QC; (d) same as (b) but for the CODC-QC; (e) location of selected profiles.

6. Application of the CODC-QC to the global hydrographic dataset

The new system was applied to the WOD18 global archive of temperature profiles. The yearly rejection rates were calculated for nine WOD instrumentation type groups (Fig. 10). The rejection rate is defined as the ratio of the number of observations flagged as bad to the total number of observations. For the entire WOD18 dataset (1940 to 2020), 9.02% (218,319,199) measurements were rejected. Among the distinct data sets the CTD, and PFL data exhibit the lowest rejection rate, being superior in accuracy and quality than the XBT and old instruments like MBT and Nansen casts (Locarnini et al., 2019). XBT data have the highest rejection rate (15.18%) because of numerous instrument-specific data quality issues (see the discussion below). The overall rejection rate

is 6.23% if not accounting for XBT data.

For each of the nine instrumental types, the local climatological range checks result in the largest outlier percentages. The variation of the rejection rates over time is evident for all instrument types. For instance, the OSD profiles are characterized by high rejection rates during the 1940s (the years of the Second World War). Here, the failure of the local bottom depth check is attributed to the higher uncertainty in profile positioning. Higher percentage of outliers for Argo profiles before 1996 corresponds to the initial stage of the float deployment with technological improvements leading to the reduction of the outlier rate afterwards. The time dependency for the XBT rejection rate is more complex. Before 1987 the rejection rate remains rather stable (8-10%), increasing since then up to 22% in 2010. We explain this by the increasing number of T7 and DB profiles which are more prone to different kind of errors compared to the shallower T4 and T6 probes, especially in the deeper ocean below 760m (Figs. 11, 12), which is beyond the nominal depth specified by manufactures but these deep layer data are always being retained in the database. The outlier percentage for the MBT data remains rather stable (5-8%) between 1940 and 1985 indicating that the instrumentation type has not undergone any significant changes. The rejection rate for the MRB exhibits strong time variability and can be mainly attributed to the malfunction of instruments anchored at several mooring sites. A high percentage of outliers for DRB profiles in 2013 is linked to the wrong geographic coordinates. Finally, for the relatively small dataset obtained by GLD (glider), the CODC-QC identifies a high outlier percentage for the profiles around Australia.

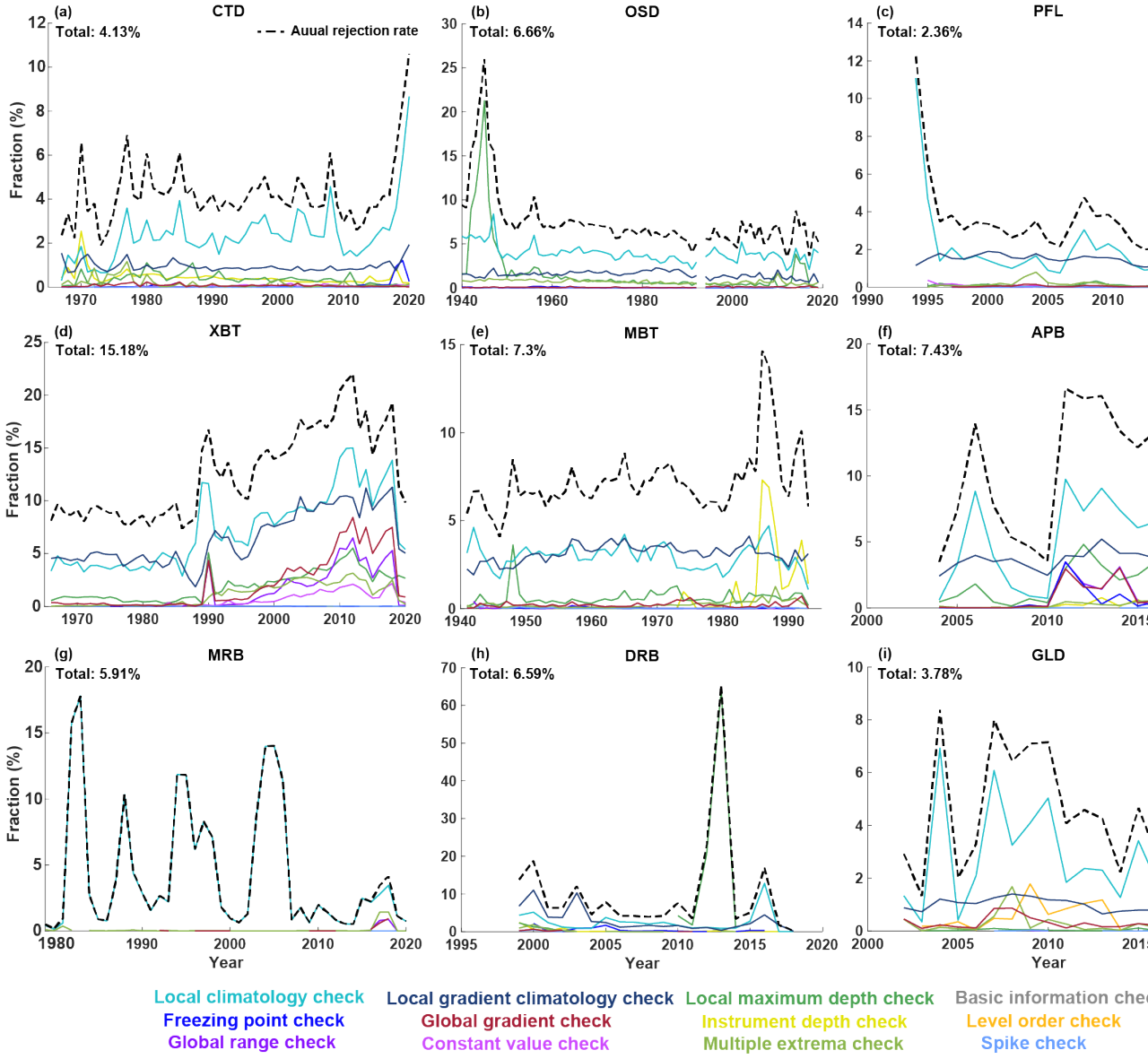

Fig. 10. Yearly percentage of data rejected by each quality checks for nine instrumentation types: CTD, OSD, PFL, XBT, MBT, APB, MRB, DRB, GLD. Black dashed line is the yearly rejection rate based on all distinct checks (this value is less than the sum of the individual check percentages because many measurements are flagged by several distinct checks in parallel. The overall rejection rate is shown at the top of each plot.

Fig. 11 shows the percentage of rejected data versus sample depth. The

CTD, PFL, and OSD data (the most accurate instrumentation types within the archive) are characterized by the rejection rate only slightly changing with depth. The exception is the CTD data from the near-surface levels shallower than 2 dbar which often correspond to the measurements registered during the sensor adjustment time before the downward cast begins. The outlier percentage change with depth is most prominent for the XBT and APB data. The sudden increase of the percentage for the XBT data is observed below the nominal maximum depth of T7 and DB probes at 760 m. Below that level the XBT acquisition systems often continue data registration, though the data become less reliable due to increased risk of the wire break, current leakage, wire stretch and bottom hit. The explanation of the increased rejection rate below 600 m for APB profiles require further investigation, one possible reason is the marine mammals barely dive into the deep layers below 600m (Boehlert et al., 2001).

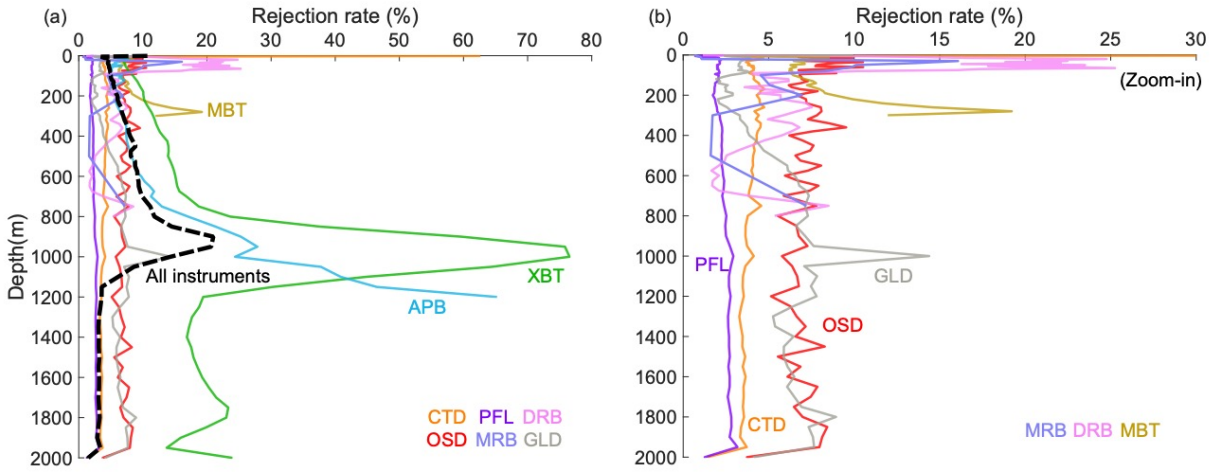

Fig. 11. (a) Percentage of outliers for different instrument types versus sample depth after applying the CODC-QC. The overall rejection rate is shown in dashed black line and different instruments in different colors. (b) is the zoomed version of (a) with XBT and APB removed.

In addition to the distributions of outlier percentage versus time and sample depth, we also provide spatial rejection rate maps for six most numerous instrumentation types: OSD, CTD, PFL, APB, MBT, and XBT (Fig. 12). The three most accurate types (OSD, CTD, and PFL) are characterized by low (less than 1%) outlier percentage for the most regions of the world ocean. A somewhat higher percentages are found within the regions with high eddy-activity like Gulf Stream, Kuroshio, Antarctic Circumpolar Current, Agulhas Return Current, Brazil current. This spatial pattern indicates the need for further adjustment of the local climatological limits for temperature and temperature gradient. The map for the XBT outlier percentage reveals a very non-homogeneous spatial pattern, with high rejection rates specific to a few routine XBT lines,

e.g., Suez channel to Australia, Tasmania to Antarctica. This provides a metric for assessing the quality of the XBT cruise lines.

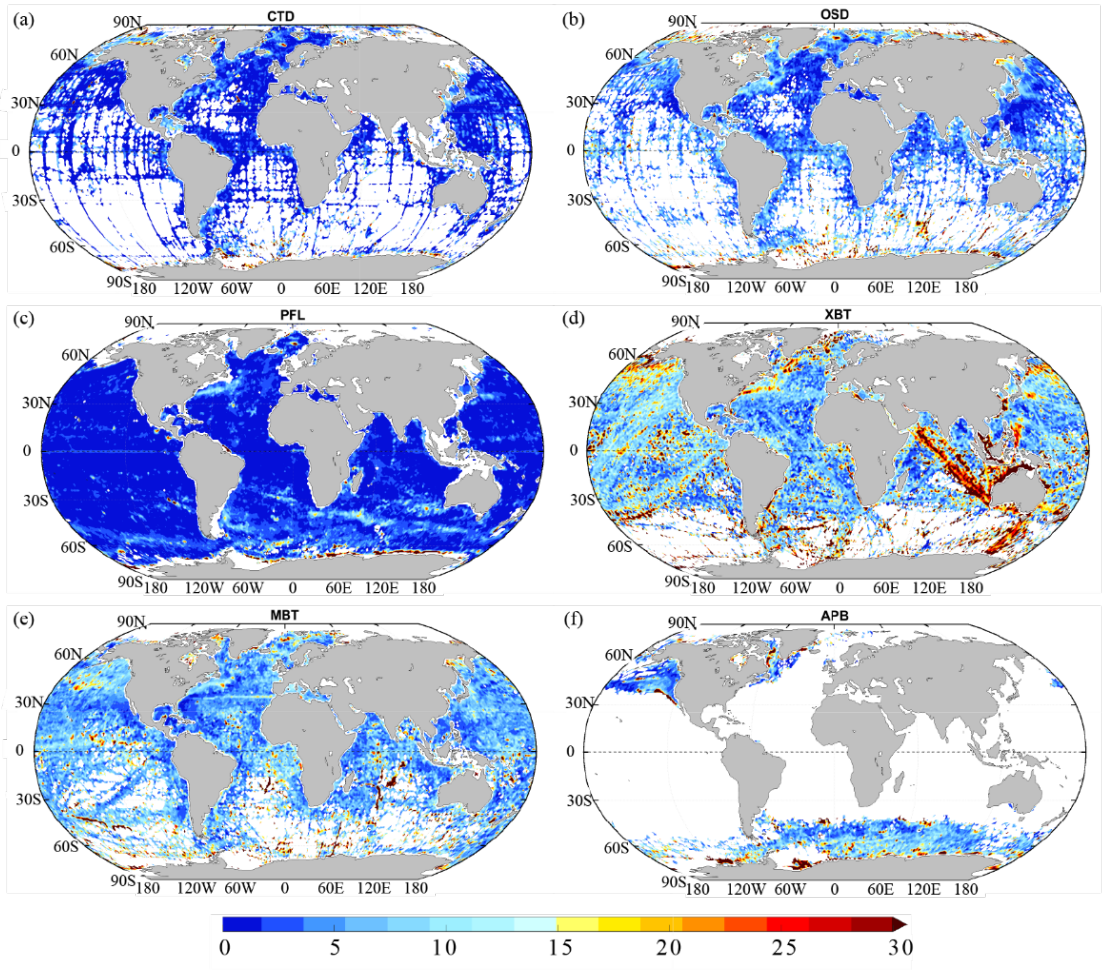

Fig. 12. The percentage of temperature outliers within 1x1-degree boxes for six main instrumentation types (CTD, OSD, PFL, MBT, XBT, APB) after applying the CODC-QC system.

Further we note that one strength of the CODC-QC is to better detect the quality issues related to XBT data, which have been raised in several studies (Cheng et al., 2016; Gouretski & Koltermann, 2007; Hanawa et al., 1995).

Specifically, these quality issues can be linked to the wire damage (e.g., wire break, wire insulation damage, wire leakage, wire stretch), problems with the launcher (e.g., false launch trigger, faulty launcher breech contact), probe, and bottom hit. Many studies developed AutoQC methods to detect these problems (Bailey et al., 1994; Daneshzadeh et al., 1994; Thadathil et al., 2001). Fig. 13 demonstrates the effectiveness of the CODC-QC for the detection of the spurious XBT profiles. Fig. 13a shows the mean and standard deviation of the temperature anomalies within 2005-2017 for the North Pacific relative to the monthly climatology. Application of the CODC-QC reduces the overall standard deviation, which decreases monotonically below the seasonal thermocline. Using the WOD-QC flag leads to the larger standard deviation and results in a spurious large magnitude of variation around 900m, that is below the terminate depth for DB and T7 probes. The composite plots of the XBT temperature profiles (Fig. 13b, c) demonstrate that the CODC-QC effectively removes spurious profiles, especially for the XBT profiles below the maximum nominal depth (~760m) of T7 and DB deep probes.

Finally, we note that the CODC-QC has rejected 9.04% data, which is 6.8 times more than the WOD-QC (1.32%) and 2.1 times more than the ICDC-QC system (4.24%). Note that both WOD-QC and ICDC-QC does not have the local temperature gradient climatological range and the XBT instrument specific check, so it is natural that this system leads to lower rejection rate.

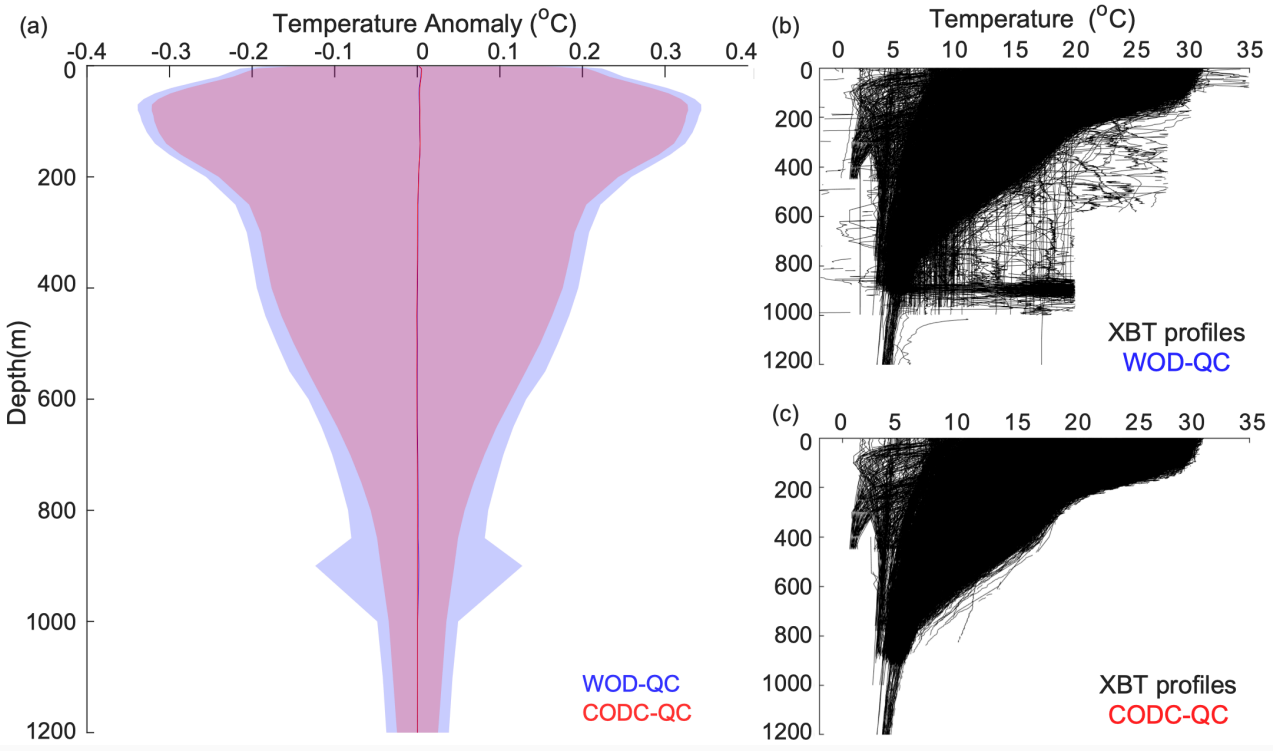

Fig. 13. (a) The north Pacific 1x1° grid-averaged monthly QCed temperature anomaly profiles (relative to 2008-2012 climatology referenced from Cheng et al. (2017)) with 1 time standard deviation envelope. (b-c) are corresponding individual XBT profiles after the application of the WOD-QC and the CODC-QC.

7. Impact of the QC procedures on the estimation of the OHC time evolution.

One of the key applications of *in-situ* ocean temperature observations is the estimation of the OHC change over time, which is a key indicator for climate change. Roemmich et al. (2019), Cowley et al. (2021) and Cheng et al. (2022) suggest that QC is one of the most important uncertainty sources contributing to the estimated OHC. Here we investigate the impact of QC on OHC by applying two AutoQC systems, including the CODC-QC and the WOD-QC. The OHC time series is calculated following the method developed by Cheng et al. (2017). The estimates of OHC are done using only the data which passed quality checks of the respective QC system.

A notable difference in OHC time series for the layer 0-2000 m are found for the two QC systems throughout the 1955-2020 period (Fig. 14a). For 1955 to 1990, the linear trend of OHC0-2000m for CODC-QC is 0.17±0.01W/m^2, being 41.7% larger than the trend estimate for WOD-QC. For 1991-2020, the warming rates are 0.64±0.02 W/m^2 and 0.61±0.02W/m^2 for CODC-QC and WOD-QC respectively, with the former being 4.9% higher. Even after the full-scale implementation of Argo floats (i.e., since about 2005) with better data quality and improved data coverage, the difference in OHC trend between CODC-QC and WOD-QC still amounts to 4.3%. These tests indicate that the choice of the QC procedure is a non-negligible source of uncertainty in OHC estimation.

Besides trends, the month-to-month variation of the rate of change of OHC, quantified by the standard deviation of the detrended OHC time series, is 17.6% smaller in CODC-QC than WOD-QC (Fig.14b), with standard deviation values of 6.09 W/m^2 for CODC-QC and 7.38 W/m^2 for WOD-QC. Trenberth et al. (2016) argued that the observed short-term OHC variability is unrealistically large because of the data noise. Therefore, the reduction of the variance when applying the new CODC-QC system implies that it has an advantage to reduce the noise of OHC and improve the estimate.

To examine the local impact of QC procedures on OHC estimates, we show the local OHC linear trend differences (OHC-WOD-QC minus OHC-CODC-QC) for three selected time periods (Fig. 15). The differences due to the QC procedures are largest for two earlier time periods: 1960-1979 and 1980-1999. But the differences are generally negative for 1960-1979 (up to –3 W m^{-2}) and positive for 1980-1999 (up to 3 W m^{-2}), implying a significant impact of QC on decadal scale OHC variation. For the latest period 200-2020 the impact of the QC procedure is much smaller (maximum of –1.5 W m^{-2}) and mostly negative over the global oceans, indicating the improved overall data quality with Argo

profiles dominating in the hydrographic archive. The pattern and signs are the differences are not fully explained in this study because it is associated with the underlying gap-filling strategy which requires further exploration.

In summary, our tests indicate that the choice of the QC procedure represents a non-negligible factor in OHC estimation: QC in different performances could potentially lead to a large difference of the OHC. However, we note that the impact of QC is investigated for only two AutoQC systems and Cheng et al. (2017) mapping method in this study, a comprehensive assessment of the QC's impact remains a research priority.

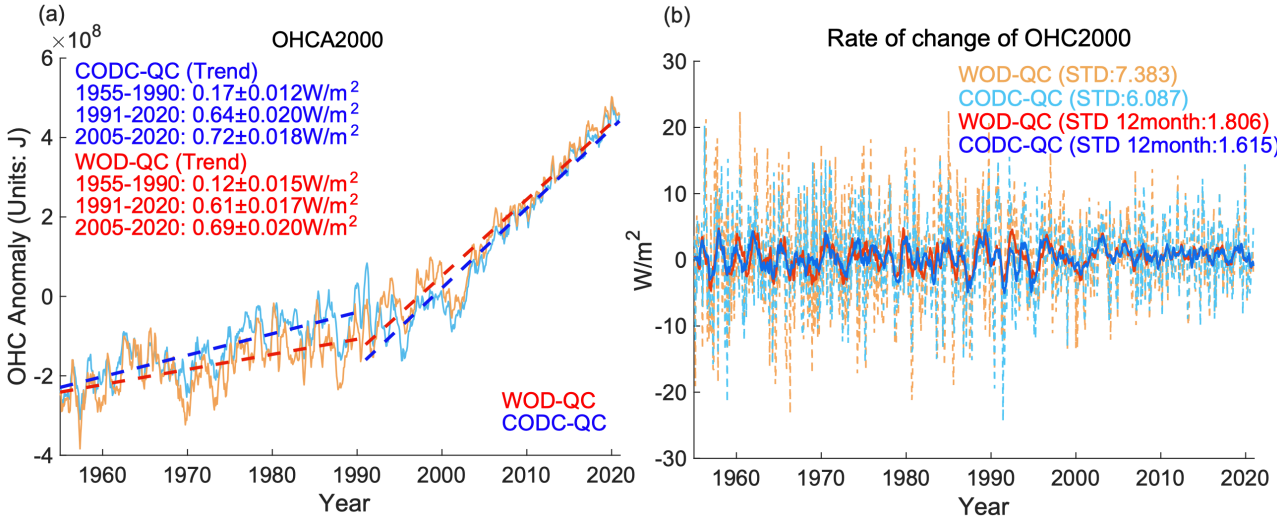

Fig. 14 (a) Global ocean heat content (OHC) anomaly (J) from 1955 to 2020 (layer 0-2000m, baseline reference climatology 1981-2010) based on WOD-QC and CODC-QC. The thick dashed lines represent the linear trends calculated as the least square fit for three time periods (1955-1990, 1991-2020, 2005-2020); (b) monthly rates (dash lines) of OHC change (J/m^2) (global area, layer 0-2000m) based on WOD-QC and CODC-QC. The rate is calculated as a simple one-sided difference as in Trenberth et al. (2016). The solid line is the 12-month running mean.

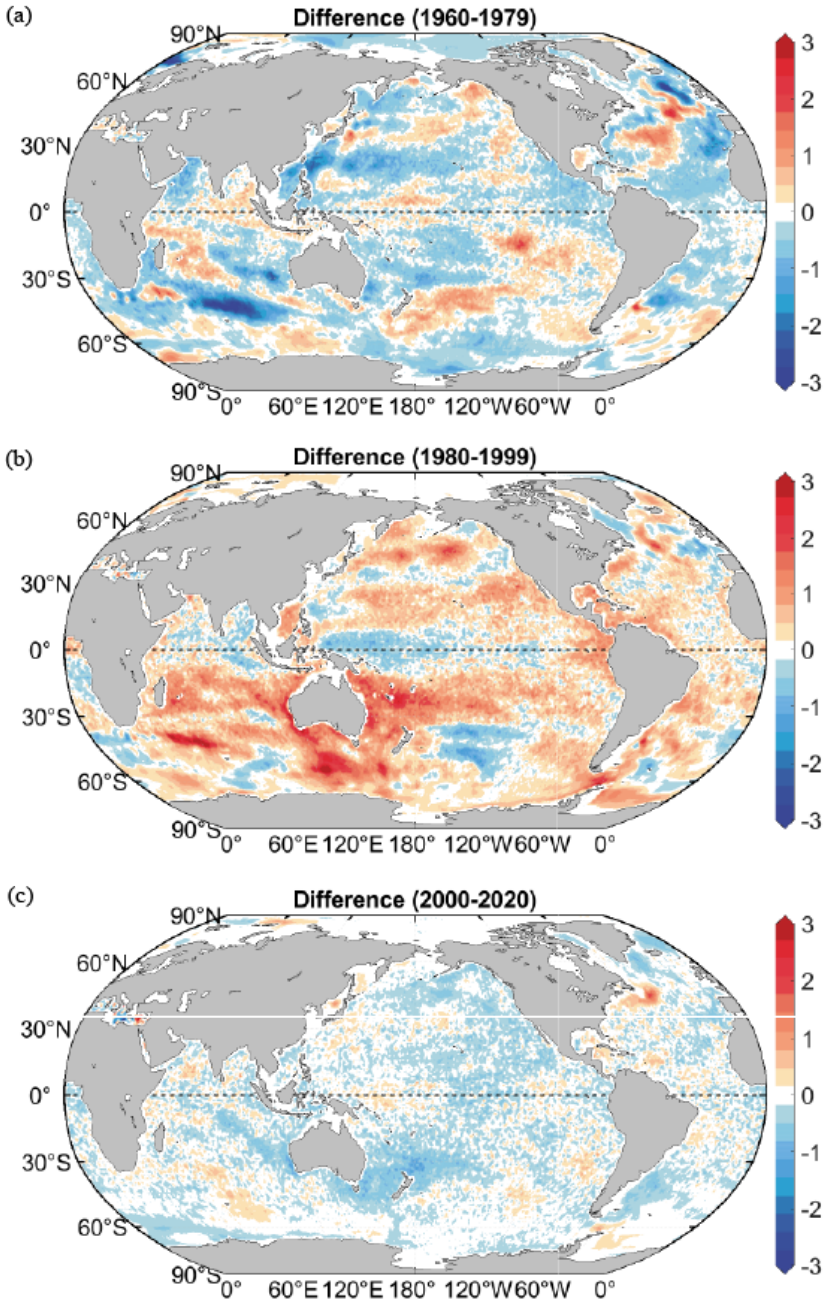
(a)
Difference (1960-1979)
(b)
Difference (1980-1999)
(c)
Difference (2000-2020)
90°N
60°N
30°N
0°
30°S
60°S
90°S
0°
60°E
120°E
180°
120°W
60°W
0°
3
2
1
0
-1
-2
-3

Fig. 15. Difference (OHC-WOD-QC minus OHC-CODC-QC) of the spatial linear trends of 0-2000m OHC (W/m^2) for three time periods (a) 1960-1979, (b) 1980-1999, and (c) 2000-2020.

8. Conclusions

In this study, we presented a new AutoQC system (CODC-QC) for the *in-situ* temperature observations. The novel feature of the QC procedure is the improved local climatological range checks, with the local temperature gradient climatological range climatology (using the IAP-TG-range field) added by the local temperature climatological range check (using the IAP-T-range field).

The performance of the CODC-QC is also evaluated using two benchmarks manual QCed datasets and compared the performance with other three AutoQC systems. We show that our new system is superior in its ability to retain good data and to flag outliers, compared to several available schemes.

The new QC procedure was applied to the global archive of temperature profiles (i.e., the WOD18). Based on the outlier statistics we find the CTD and Argo float data being superior in quality followed by the old Nansen cast, MBT and APB instrumentation groups. XBTs represent the instrumentation group with the highest percentage of outliers due to the numerous instrument specific quality issues.

Finally, we estimated the impact of QC on the estimation of global and regional OHC change by testing CODC-QC and WOD-QC procedures. We found that the difference between the calculated linear OHC trend amounts to 41.67%, 4.92% and 4.34% for the time periods 1955 to 1990, 1991 to 2020 and 2005 to 2020 respectively. Application of the more stringent CODC-QC procedure leads to the increased warming trend. In addition, the CODC-QC reduces the month-to-month variation of the global ocean warming rate estimate by 17.55%.

It is worthwhile to discuss some potential improvements on QC system, which deserves further studies. For example: (1) one task of QC system is to identify spurious profiles that break the physical laws, this means that the vertical structure of temperature profiles should be better defined and used in QC system. Currently, only vertical gradient information is used because of the difficulty in defining the "shape" of a profile. A possible solution is that the application of machine learning approach (i.e., artificial intelligence QC), which is a non-parameterized approach and does not require an explicit definition of the 'profile shape'. This could be a useful addition to the traditional AutoQC (e.g., Mieruch et al. (2021)). (2) The extreme events should be treated specifically in QC system in the future, for examples, the marine heat waves and tropical cyclones will lead to a dramatic change of ocean conditions, which is likely to be flagged by most of the QC systems including the CODC-QC. Thus, cause is needed when applying the QC flags to study these extreme events.

9. Acknowledgements

This study is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences [Grant no. XDB42040402], the National Natural Science Foundation of China [Grant no. 42122046, 42076202), National Key R&D Program of China [Grant no. 2017YFA0603202], Key Deployment Project of Centre for Ocean Mega-Research of Science, CAS [Grant no. COMS2019Q01] and open fund of State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, MNR [Grant no. QNHX2133]. We thank the help of the IQuOD team for providing the open-source AutoQC packages and Oceanographic Data Center, CAS for support in data analysis.

10. Data/Code Available Statement

All the data we used here are freely accessible. The WOD18 temperature data can be easily downloaded in the NCEI/NOAA data access pool (https://www.ncei.noaa.gov/access/world-ocean-database-select/dbsearch.html). The QuOTA dataset and WOCE one-time profile dataset can be freely downloaded via CSIRO data pool (https://doi.org/10.25919/5ec357563bd3e) and the U.S. National Ocean Data Center (https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:NODC-WOCE-GDR), respectively. The code of CODC-QC is available as an Open-Source Python package that is freely available from Github (https://github.com/zqtzt/CODCQC) under the Apache-2.0 License. The average run time to quality control a typical temperature profile is ~0.08s. More information of the algorithm and data reported in this study (including climatologies development and CODC-QCed data) are available via http://www.ocean.iap.ac.cn/.

References

https://doi.org/10.1029/2003GL018851

http://dx.doi.org/10.17882/42182

https://doi.org/10.1175/1520-0426(2001)018

https://doi.org/10.1016/j.jmarsys.2008.12.015

https://doi.org/10.13155/78994

https://doi.org/10.1175/Bams-D-15-00031.1

https://doi.org/10.1007/s00376-022-1461-3

https://doi.org/10.1126/sciadv.1601545

https://doi.org/10.1175/JCLI-D-20-0366.1

https://doi.org/10.1175/Jtech-D-13-00197.1

https://doi.org/10.3389/fmars.2021.689695

https://doi.org/10.3389/fmars.2021.689695

https://doi.org/10.3389/fmars.2019.00452

https://doi.org/10.1002/2013jc009067

https://doi.org/10.5194/os-14-1127-2018

https://doi.org/10.1080/16742834.2019.1588066

https://doi.org/10.1175/JTECH-D-19-0205.1

https://doi.org/10.1016/S0079-6611(00)00049-5

https://doi.org/10.1029/2006gl027834

https://doi.org/10.1175/JTECHO539.1

https://doi.org/10.1016/0967-0637(95)97154-Z

https://doi.org/10.1007/s13131-017-1035-x

https://doi.org/10.1175/1520-0485(1983)013

https://doi.org/10.3389/fmars.2021.611742

https://doi.org/10.1126/science.166.3903.373

https://doi.org/10.1029/JC082i009p01369

https://doi.org/10.3389/fmars.2019.00439

https://doi.org/10.1002/jgrc.20122

https://doi.org/10.3389/fenvs.2021.711363

https://doi.org/10.3389/fmars.2019.00444

https://doi.org/10.1029/2019EA001019

https://doi.org/10.1007/s11430-021-9846-7

https://doi.org/10.1016/S0098-3004(00)00172-2

https://doi.org/10.1029/2019EA000658

https://doi.org/10.1175/jcli-d-16-0339.1

http://doi.org/10.7289/V52N50GF

https://doi.org/10.1175/1520-0442(2000)013

https://doi.org/10.1175/1520-0442(2000)013

https://doi.org/10.1175/2008jcli2290.1

http://dx.doi.org/10.13155/33951

https://doi.org/10.2307/4586294

Adil, I. H., & Irshad, A. R. (2015). A modified approach for detection of outliers. *Pakistan Journal of Statistics and Operation Research, 11*, 91-102. Antonov, J. I., Levitus, S., & Boyer, T. P. (2004). Climatological annual cycle of ocean heat content. *Geophysical Research Letters, 31*(4), 1-4. Argo. (2000). Argo

float data and metadata from global data assembly centre. *Sea Scientific Open Data Edition, Argo Global Data Assembly Centre (GDAC)*. Bailey, R., Gronell, A., Phillips, H., Tanner, E., & Meyers, G. (1994). Quality control cookbook for XBT data, Version 1.1. *CSIRO Marine Laboratories Reports, 221*. Barton, Z., & Gonzalez, I. (2016). *AOML high density XBT system setup instructions and troubleshooting manual*. NOAA manual: Boehlert, G. W., Costa, D. P., Crocker, D. E., Green, P., O'Brien, T., Levitus, S., & Le Boeuf, B. J. (2001). Autonomous pinniped environmental samplers: Using instrumented animals as oceanographic data collectors. *Journal of Atmospheric and Oceanic Technology, 18*(11), 1882-1893. <1882:APESUI>2.0.CO;2Boyer, T. P., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., Locarnini, R. A., Mishonov, A. V., Paver, C. R., Reagan, J. R., Seidov, D., Smolyar, I. V., Weathers, K., & Zweng, M. M. (2019). World Ocean Database 2018. *A. V. Mishonov, Technical Editor, NOAA Atlas NESDIS 87*. Brander, K. (2010). Impacts of climate change on fisheries. *Journal of Marine Systems, 79*(3-4), 389-402. Cabanes, C., Angel-Benavides, I., Buck, J., Coatanoan, C., Dobler, D., Herbert, G., Klein, B., Maze, G., Notarstefano, G., & Owens, B. (2021). DMQC cookbook for core Argo parameters. Catherine, S., Herve, C., Antoine, P., & Fabrizio, D. O. (2014). *Bio-Argo quality control manual for the Chlorophyll-A concentration. IFREMER* Cheng, L., Abraham, J., Goni, G., Boyer, T., Wijffels, S., Cowley, R., Gouretski, V., Reseghetti, F., Kizu, S., Dong, S. F., Bringas, F., Goes, M., Houpert, L., Sprintall, J., & Zhu, J. (2016). XBT science: assessment of instrumental biases and errors. *Bulletin of The American Meteorological Society, 97*(6), 923-934. Cheng, L., Abraham, J., Trenberth, K. E., Fasullo, J., Boyer, T., Mann, M. E., Zhu, J., Wang, F., Locarnini, R., Li, Y., Zhang, B., Tan, Z., Yu, F., Wan, L., Chen, X., Song, X., Liu, Y., Reseghetti, F., Simoncelli, S., Gouretski, V., Chen, G., Mishonov, A., & Reagan, J. (2022). Another record: Ocean warming continues through 2021 Despite La Niña Conditions. *Advances In Atmospheric Sciences*. Cheng, L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., & Zhu, J. (2017). Improved estimates of ocean heat content from 1960 to 2015. *Science Advances, 3*(3), e1601545. Cheng, L., Trenberth, K. E., Gruber, N., Abraham, J. P., Fasullo, J. T., Li, G., Mann, M. E., Zhao, X., & Zhu, J. (2020). Improved estimates of changes in upper ocean salinity and the hydrological cycle. *Journal of Climate, 33*(23), 10357-10381. Cheng, L., Zhu, J., Cowley, R., Boyer, T., & Wijffels, S. (2014). Time, probe type, and temperature variable bias corrections to historical expendable bathythermograph observations. *Journal of Atmospheric and Oceanic Technology, 31*(8), 1793-1825. Cowley, R., Killick, R. E., Boyer, T., Gouretski, V., Reseghetti, F., Kizu, S., Palmer, M. D., Cheng, L., Storto, A., Le Menn, M., Simoncelli, S., Macdonald, A. M., & Domingues, C. M. (2021). International Quality-Controlled Ocean Database (IQuOD) v0.1: the temperature uncertainty specification. *Frontiers in Marine Science, 8*(607), 689-695. Daneshzadeh, Y. H., Festa, J. F., & Minton, S. M. (1994). *Procedures used at AOML to quality control real time XBT data collected in the Atlantic Ocean*. Miami, Florida: Atlantic Oceanographic Meteorological Laboratory.Domingues, C., & Palmer, M. (2015). The IQuOD initiative: towards an international quality controlled ocean database. *Frontiers in Marine*

Science, 67, 38-40. Garcia, H. E., Boyer, T. P., Locarnini, R. A., Baranova, O. K., & Zweng, M. M. (2018). World Ocean Database 2018: User's Manual. *A. V. Mishonov, Technical Ed., NOAA, Silver Spring, MD.* Goni, G., Sprintall, J., Bringas, F., Cheng, L., Cirano, M., Dong, S., Domingues, R., Goes, M., Lopez, H., Morrow, R., Rivero, U., Rossby, T., Todd, R. E., Trinanes, J., Zilberman, N., Baringer, M., Boyer, T., Cowley, R., Domingues, C. M., Hutchinson, K., Kramp, M., Mata, M. M., Reseghetti, F., Sun, C., Bhaskar, U., & Volkov, D. (2019). More than 50 years of successful continuous temperature section measurements by the global expendable bathythermograph network, its integrability, societal benefits, and future. *Frontiers in Marine Science, 6*, 452. Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans, 118*(12), 6704-6716. Gouretski, V. (2018). World Ocean Circulation Experiment – Argo global hydrographic climatology. *Ocean Science, 14*(5), 1127-1146. Gouretski, V. (2019). A new global ocean hydrographic climatology. *Atmospheric and Oceanic Science Letters, 12*(3), 226-229. Gouretski, V., & Cheng, L. (2020). Correction for systematic errors in the global data set of temperature profiles from mechanical bathythermographs. *Journal of Atmospheric and Oceanic Technology, 37*(5), 841-855. Gouretski, V., & Jancke., K. (2001). Systematic errors as the cause for an apparent deep water property variability: Global analysis of the WOCE and historical hydrographicdata. *Progress In Oceanography, 48*, 337–402. Gouretski, V., & Koltermann, K. (2007). How much is the ocean really warming? *Geophysical Research Letters, 34*(1), 1-5. Gouretski, V., & Koltermann, K. P. (2004). WOCE global hydrographic climatology. *Berichte des BSH, 35*, 1-52. Gronell, A., & Wijffels, S. E. (2008). A semiautomated approach for quality controlling large historical ocean temperature archives. *Journal of Atmospheric and Oceanic Technology, 25*(6), 990-1003. Hanawa, K., Rual, P., Bailey, R., Sy, A., & Szabados, M. (1995). A new depth-time equation for Sippican or TSK T-7, T-6 and T-4 expendable bathythermographs (XBT). *Deep Sea Research Part I: Oceanographic Research Papers, 42*(8), 1423-1451. Hubert, M., & Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis, 52*, 5186-5201. Intergovernmental Oceanographic Commission. (2010). GTSPP real-time quality control manual. Revised Edition 2010. UNESCO. IPCC. (2021). The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press. In Press.* King, B. A., Firing, E., & Joyce, T. M. (2001). *Shipboard observations during WOCE* (Vol. 77). International Geophysics: Elsevier.Levitus, S. (1982). *Climatological atlas of the world ocean, Volume 13, US Department of Commerce, National Oceanic and Atmospheric Administration* Levitus, S., & Boyer, T. P. (1994). *World ocean atlas 1994. volume 4. temperature, No. PB-95-270112/XAB; NESDIS-4. National Environmental Satellite, Data, and Information Service, Washington, DC (United States)* Liu, Z., Wu, X., Xu, J., Li, H., Lu, S., Sun, C., & Cao, M. (2017). China Argo project: progress in China Argo ocean observations and data applications. *Acta Oceanologica Sinica, 36*(06),

1-11. Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., Reagan, J. R., Seidov, D., Weathers, K. W., Paver, C. R., & Smolyar, I. V. (2019). World Ocean Atlas 2018, Volume 1: Temperature. A. Mishonov, Technical Editor. NOAA Atlas NESDIS 81. *52pp.* Luyten, J., Pedlosky, J., & Stommel, H. (1983). The ventilated thermocline. *Journal of Physical Oceanography, 13*(2), 292-309. <0292:TVT>2.0.CO;2Mieruch, S., Demirel, S., Simoncelli, S., Schlitzer, R., & Seitz, S. (2021). SalaciaML: A deep learning approach for supporting ocean data quality control. *Frontiers in Marine Science, 8*(314), 611-742. Neal, V. T., Neshyba, S., & Denner, W. (1969). Thermal stratification in the Arctic Ocean. *Science, 166*(3903), 373-374. Oakey, N., & Elliott, J. (1977). Vertical temperature gradient structure across the Gulf Stream. *Journal of Geophysical Research, 82*(9), 1369-1380. Ono, S., Matsuyama, H., Fukui, K., & Hosoda, S. (2015). *A preliminary study on quality control of oceanic observation data by machine learning methods.* Paper presented at the the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems, Singapore.Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., Oke, P., Owens, W. B., Pouliquen, S., Purkey, S., Scanderbeg, M., Suga, T., Wijffels, S., Zilberman, N., Bakker, D., Baringer, M., Belbeoch, M., Bittig, H. C., Boss, E., Calil, P., Carse, F., Carval, T., Chai, F., Conchubhair, D. Ó., d'Ortenzio, F., Dall'Olmo, G., Desbruyeres, D., Fennel, K., Fer, I., Ferrari, R., Forget, G., Freeland, H., Fujiki, T., Gehlen, M., Greenan, B., Hallberg, R., Hibiya, T., Hosoda, S., Jayne, S., Jochum, M., Johnson, G. C., Kang, K., Kolodziejczyk, N., Körtzinger, A., Traon, P. Y. L., Lenn, Y. D., Maze, G., Mork, K. A., Morris, T., Nagai, T., Nash, J., Garabato, A. N., Olsen, A., Pattabhi, R. R., Prakash, S., Riser, S., Schmechtig, C., Schmid, C., Shroyer, E., Sterl, A., Sutton, P., Talley, L., Tanhua, T., Thierry, V., Thomalla, S., Toole, J., Troisi, A., Trull, T. W., Turton, J., Velez-Belchi, P. J., Walczowski, W., Wang, H., Wanninkhof, R., Waterhouse, A. F., Waterman, S., Watson, A., Wilson, C., Wong, A. P. S., Xu, J., & Yasuda, I. (2019). On the future of Argo: A global, full-depth, multi-disciplinary array. *Frontiers in Marine Science, 6*(439), 439. Schmidtko, S., Johnson, G. C., & Lyman, J. M. (2013). MIMOC: A global monthly isopycnal upper-ocean climatology with mixed layers. *Journal of Geophysical Research: Oceans, 118*(4), 1658-1672. Shahzadi, K., Pinardi, N., Barth, A., Troupin, C., Lyubartsev, V., & Simoncelli, S. (2021). A new global ocean climatology. *Frontiers in Environmental Science, 9*(363), 711363. Stammer, D., Bracco, A., AchutaRao, K., Beal, L., Bindoff, N. L., Braconnot, P., Cai, W., Chen, D., Collins, M., Danabasoglu, G., Dewitte, B., Farneti, R., Fox-Kemper, B., Fyfe, J., Griffies, S. M., Jayne, S. R., Lazar, A., Lengaigne, M., Lin, X., Marsland, S., Minobe, S., Monteiro, P. M. S., Robinson, W., Roxy, M. K., Rykaczewski, R. R., Speich, S., Smith, I. J., Solomon, A., Storto, A., Takahashi, K., Toniazzo, T., & Vialard, J. (2019). Ocean climate observing requirements in support of climate research and climate information. *Frontiers in Marine Science, 6*, 444. Sugiura, N., & Hosoda, S. (2020). Machine learning technique using the signature method for automated quality control of Argo profiles. *Earth and Space Science, 7*(9), e2019EA001019. Tan, Z., Zhang, B., Wu, X., Dong, M., & Cheng, L. (2022). Quality control for ocean observa-

tions: From present to future. *Science China Earth Sciences, 65*(2), 215-233. Thadathil, P., Ghosh, A. K., Sarupria, J., & Gopalakrishna, V. (2001). An interactive graphical system for XBT data quality control and visualization. *Computers & Geosciences, 27*(7), 867-876. Tozer, B., Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., & Wessel, P. (2019). Global bathymetry and topography at 15 arc sec: SRTM15+. *Earth and Space Science, 6*. Trenberth, K., Fasullo, J., von Schuckmann, K., & Cheng, L. (2016). Insights into Earth's energy imbalance from multiple sources. *Journal of Climate, 29*(20), 7495-7505. U.S. Integrated Ocean Observing System. (2015). Manual for real-time quality control of dissolved nutrients data: A guide to quality control and quality assurance of coastal and dissolved nutrients observations. Vanderviere, E., & Huber, M. (2004). An adjusted boxplot for skewed distributions. *COMPSTAT'2004 Symposium, 23-27*, 1933-1940. <0596:AAOTTI>2.0.CO;2Wang, B., Wu, R., & Lukas, R. (2000). Annual adjustment of the thermocline in the tropical Pacific Ocean. *Journal of Climate, 13*(3), 596-616. <0596:AAOTTI>2.0.CO;2WCRP. (1988a). *World Ocean Circulation Experiment implementation plan I. Detailed requirements. WCRP #11, WMO/TD #242, WOCE report no. 20/88* WCRP. (1988b). *World Ocean Circulation Experiment implementation plan II. Scientific background. WCRP #12, WMO/TD #243, WOCE report no. 21/88* Wijffels, S., Willis, J., Domingues, C., Barker, P., White, N. J., Gronell, A., Ridgway, K., & Church, J. A. (2008). Changing expendable bathythermograph fall rates and their impact on estimates of thermosteric sea level rise. *Journal of Climate, 21*(21), 5657-5672. Wong, A., Keeley, R., & Carval, T. (2020). Argo quality control manual for CTD and trajectory data. Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis with special reference to x-ray techniques. *Public Health Reports, 62*(2), 1432-1439.