A Quality Control Procedure for Climatological Studies Using Argo Data in the North Pacific Western Boundary Current Region

WENJING JIA AND DONG WANG

School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing, China, and Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

NADIA PINARDI

Istituto Nazionale di Geofisica e Vulcanologia, and Department of Physics and Astronomy, University of Bologna, Bologna, Italy

SIMONA SIMONCELLI

Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy

ANDREA STORTO

Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

SIMONA MASINA

Centro Euro-Mediterraneo sui Cambiamenti Climatici, and Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy

(Manuscript received 14 July 2015, in final form 28 March 2016)

ABSTRACT

A quality control (QC) procedure is developed to estimate monthly mean climatologies from the large Argo dataset (2005–12) over the North Pacific western boundary current region. In addition to the individual QC procedure, which checks for instrumental, transmission, and gross errors, the paper describes and shows the impact of climatological checks (collective QC) on the quality of both processed profiles and resultant climatological distributions. Objective analysis (OA) is applied progressively to produce the gridded climatological fields. The method uses horizontal regional climatological averages defined in five regime-oriented subregions in the Kuroshio area and the Japan Sea. Performing the QC procedure on specific coherent subregions produces improved profiling data and climatological fields because more details about the local hydrodynamics are taken into consideration. Nonrepresentative data and random noises are more effectively rejected by this method, which has value both in defining a climatological mean and identifying outlier data. Assessing with both profiling and coordinated datasets, the agreement is reasonably good (particularly for those areas with abundant observations), but the results (although already smoothed) can capture more detailed or mesoscale features for further regional studies. The method described has the potential to meet future challenges in processing accumulating Argo observations in the coming decades.

1. Introduction

Featured by mesoscale activities and strong carbon uptake (e.g., Takahashi et al. 2002; Yu and Weller 2007), the western boundary current (WBC) region plays a key role in the ocean heat transport and overturning circulation. Being characterized by a frontal structure and by mesoscale and ring dynamics, the WBC and its associated recirculation subregions are challenging in terms of observational and modeling requirements for climatological studies. Even with coordinated in situ and satellite observations, it is difficult to depict the three-dimensional WBC synoptic structure (Argo Science Team 2012) because of the limitations of temporal and spatial coverage of the measurements.

Argo has been widely deployed since the 2000s and provides for the first time the capability to monitor the

Corresponding author e-mail: Dong Wang, dong.wang@nuist.edu.cn

DOI: 10.1175/JTECH-D-15-0140.1

upper ~2000-m oceanographic properties at both continuous spatial and temporal scales in a cost-efficient and versatile way. The Argo observing system has the potential to extend measurement to even deeper depths (~6000 m) and to achieve real global observations from the deep oceans to the marginal seas and the WBC region, etc., with various sampling densities according to their own characteristics and needs in the specific areas (Argo Science Team 2012). Argo data, with a more uniform sampling than other types of observations (e.g., CTD, XBT, bottle, etc.), could serve to better estimate climatologies for different parts of the World Ocean. Thus, our work concentrates on the estimation of climatologies only from the emerging and ever-growing Argo profiles. It is prospective to develop long-term multiscale three-dimensional hydrographic products using the Argo observational data for both scientific and operational uses. Argo observations have experienced technical problems from the beginning: a pressure sensor drift was detected in 2009 (Argo Steering Team 2011; Barker et al. 2011), a salinity offset due to biofouling was found to affect floats with long lifetimes in different World Ocean regions (Oka 2005; Wong et al. 2003; Böhme and Send 2005; Owens and Wong 2009), and data transmission errors were documented (Boyer et al. 2013). Data quality control (QC) is thus necessary to be used to abate or solve most of the problems. For example, Barker et al. (2011) detected and analyzed the pressure drift; salinity troubles have been calibrated by Wong et al. (2003) for the open tropical and subtropical oceans, Böhme and Send (2005) for the polar regions, and Owens and Wong (2009) for the global oceans in general. The international Argo program recommends quality control for Argo data being done in two steps: 1) real-time QC by automatic screening of errors and spikes, etc., and data transmission within 24 h to Argo Global Data Assembly Centers (GDACs); and 2) delayed-mode QC (DMQC) with more sophisticated procedures and data transmitted to GDACs every 1–2 years (Barker et al. 2011).

It is particularly difficult to implement DMQC in practice in the WBC region mainly due to the regionally intrinsic large gradient of hydrographic variables and the lack of a high-quality historical reference database for the Argo data DMQC, especially for the Kuroshio region (Argo Steering Team 2013). It is not surprising that very limited studies exist on Argo data QC designed specifically for the WBC region, and most of the previous QC studies focused merely on the large-scale open oceans—for example, the Atlantic (Gaillard et al. 2009), the Pacific (Zhang et al. 2013), the equatorial (Wong et al. 2003) and polar regions (Böhme and Send 2005)—or on the global ocean in general (e.g., Owens and Wong 2009; Johnson et al. 2013). To the best of our knowledge, this study presents the first effort to develop a QC method applied specifically to the large Argo datasets in the Kuroshio region.

In this study, we develop a QC procedure for the Argo data in the WBC's subregions and apply the method in the Kuroshio region as an example. The QC method is performed on specific coherent subregions to take more details about the regional dynamics into account than the traditional QC from the GDACs (Argo Data Management Team 2012). To facilitate the assessment of the Argo observing system in the complicated Kuroshio region, objective analysis (OA; Bretherton et al. 1976) will be applied to map profiling data onto a regular grid. Three-dimensional hydrographic climatologies in the WBC region could then be constructed. The OA-inferred distributions have the potential to support numerical model prediction and a data assimilation system (Carter and Robinson 1987), and to improve our understanding of frontal structures, turbulences and cross-frontal exchanges, and ventilation processes in such highly energetic regions.

This paper describes our efforts to elaborate the methods of processing and gridding the Argo profiling data over the WBC (Kuroshio) region in an assessment of other available approaches and datasets as well. In a larger context, we attempt to provide guidance for preprocessing of profiling data before using the data in a specific study, though the detailed techniques can be varied according to the study's purpose. The data and methods for a comprehensive examination of the Kuroshio regional Argo observing system are presented in section 2, followed by an analysis of the resulting hydrographic profiles and gridded fields in section 3. Section 4 evaluates and discusses the results in terms of both Argo profiles and gridded climatologies, and section 5 concludes the study.

2. Data and methods

a. *Observational data and study region*

The Argo data used in this study are from the real-time profiling floats during the years 2005–12 in the study domain (21°–42°N, 115°–145°E; see Fig. 1) and are accessed from the portal of the Coriolis Data Centre (http://www.coriolis.eu.org). The Argo sampling density in this 8-yr period allows us to build the monthly climatological hydrographic fields at a relatively fine horizontal resolution over the broad geographical extent of the study region. Within the temporal and spatial scopes of this study, the Argo floats, deployed primarily by the Chinese, Japanese, and Korean national programs under the international Argo program, a part of the Global

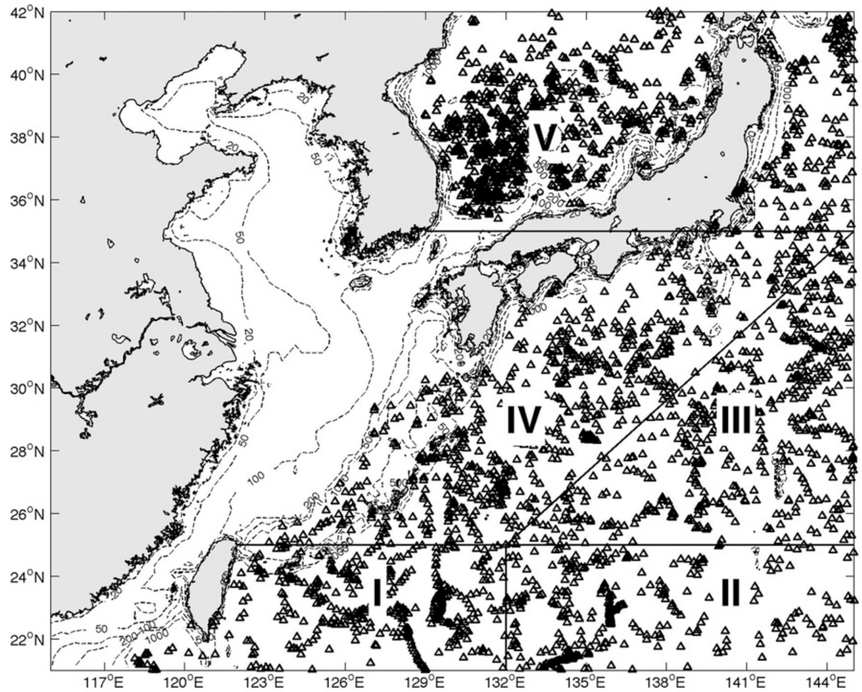

FIG. 1. Study domain (21°–42°N, 115°–145°E) with selected bathymetry contours (20, 50, 100, 200, 500, and 1000 m; dashed lines), and an example (in August) of the monthly position of Argo profiles (triangles) during the period 2005–12, accompanied by subregions (I–V, separated by black straight lines).

Ocean Observing System, are of various types (APEX, PROVOR, ARVOR, etc.) and data communication technologies (e.g., Argos, Iridium; Argo Steering Team 2013). For data processing and analysis, we include all the Argo profiles in every month of the 8-yr period over the study region, aiming at a wider end use of the Argo observing system than technical consideration of each single float. Figure 1 shows the position of floating profiles in August for the considered 8-yr period as an example, and the other 11 months have similar distributions with a sufficient number of sampling profiles to analyze and process. Topography data (partly shown in Fig. 1) are from the General Bathymetric Chart of the Oceans (GEBCO) (http://www.gebco.net/data_and_products/gridded_bathymetry_data). The desired grid (usually coarser than 1 min) of bathymetry for OA can be extracted from the GEBCO One Minute Grid without interpolation.

For the study region, we enlarge the targeted Kuroshio region, extending it eastward slightly to parts of the Kuroshio Extension (KE) and the North Pacific Subtropical Countercurrent (STCC) regions (subregions I–IV in Fig. 1), and embrace the southern area of the Japan Sea (subregion V in Fig. 1), which is connected to the Kuroshio region. In doing so we are able to obtain adequate Argo data covering a complex study region that extends from the marginal sea to the open ocean and includes the WBC transition zone, though few Argo observations exist in the Kuroshio-related East China Sea (ECS, the data-blank marginal sea areas above the continental shelf in Fig. 1). Apart from the data-blank ECS, we divide the study domain into five subregions (I–V) for further analysis, based on the dynamics of major currents and topographic and climatic effects, as illustrated in Fig. 1. Subregions I and II are the aforementioned STCC divisions; in particular, subregion II encompasses part of the North Pacific Northern Subtropical Front (NSTF). Subregion III contains the North Pacific Subtropical Mode Water (STMW) segments advected by the Kuroshio recirculation (Kobashi et al. 2006). Subregion IV is the core Kuroshio subregion. Subregions I–IV make up the main region of interest for this study and are collectively defined as the Kuroshio region. Subregion V is the southern part of the Japan Sea. Besides the five subregions, the northeastern minor portion of the study domain (in Fig. 1, northward of 35°N and east of Japan in the open North Pacific) is divided separately for Argo data processing and analysis, but it is of least concern in this study and is not assigned a subregion.

The processed Argo data will be evaluated in section 4a with those data obtained from other QC-processed profiling databases: the GDACs and the World Ocean

Database 2013 (WOD13; Boyer et al. 2013). For the same temporal and spatial scales, the resultant climatological fields from our method will be assessed (in section 4b) with two other newly developed coordinated datasets: the *World Ocean Atlas 2013* (*WOA13*; Locarnini et al. 2013; Zweng et al. 2013) and a state-of-the-art global ocean reanalysis (RA). *WOA13* is the most updated observational climatology of 2005–12 (¼° grid), derived from WOD13 of multitype in situ measurements (including CTD, XBT, Argo floats, etc.) by Locarnini et al. (2013) and Zweng et al. (2013). RA is the 2005–12 time average of a reanalyzed climatology from a long-term (1979–2012) eddy-permitting (¼° grid) global ocean reanalysis (¼° grid) that has assimilated both in situ (including Argo observations) and satellite observations (Storto and Masina 2014), and is available online via the portal of the European Commission–funded project MyOcean2.

b. QC and OA methods

The QC developed in this paper consists mainly of two steps: the first is a simple individual check, while the second step is a climatological check (or "collective QC"). During the first step, each Argo datum is checked for date, time, and position duplication; nonmonotonic pressure growth with depth; negative values, gross range, and spikes and gradients of temperature and salinity, as recommended by the Argo QC manual (Argo Data Management Team 2012). This first step is called the "individual QC" here, since each Argo profile can be examined individually, in contrast to the second step of the climatological check, which usually requires a large set(s) of profiles (i.e., collective QC). The second step is named collective QC also because it often includes iterative steps (along with OA, for spatial continuity or consistency examination of neighboring data), which is quite different from the first step in both methodology and purpose.

In the second step, a climatological check (or collective QC) is carried out for vertically bin-averaged data at standard levels (Table 1). On the basis of the normal (Gaussian) law and the Argo sampling density, the monthly climatologies of temperature and salinity are adequate for a statistical check. This check is done by flagging each datum that is outside two standard deviations from a climatological mean that is computed from the previous step's quality-checked data. The collective QC method uses the data that passed the preceding QC step to generate an OA regularly gridded climatology. Vertical climatological means and standard deviations at standard depths (Table 1) are computed by averaging the gridded climatology in each individual subregion in Fig. 1. These climatological means and standard deviations are used in the climatological check on the data that passed the preceding QC, flagging more data. Out of this process the remaining data are used again to recalculate a climatology and evaluate the mean and standard deviations, iteratively. The method progressively flags data from new climatological means evaluated without the previously flagged data after the statistical check.

This is the key step of this study to more effectively inspect nonrepresentative data and random noises. The climatological check can be achieved with either regular grids or irregular areas (Manzella and Gambetta 2013). Hence, we could find the appropriate method for the monthly climatological check through the sensitivity test of different QC approaches. The sensitivity of the collective QC method to different horizontal averaging subregions and the number of iterations will be described in section 3.

The OA technique is used to grid the monthly quality-controlled data at each standard level to check the spatial continuity or consistency of adjacent points and to produce the climatological fields at the final stage. OA is effective at checking the "bull's-eyes," the large-scale gradients over a relatively small spatial scale, particularly in the undersampled areas with large

TABLE 1. Vertical bin-averaged depth levels, based on vertical sampling densities of the Argo data, which are consistent with the standard levels of the World Ocean Database 2009 (Boyer et al. 2009) and the thickness of each level.

Level	Depth (m)	Thickness (m)
1	Surface	5
2	10	10
3	20	10
4	30	10
5	50	30
6	75	30
7	100	30
8	125	30
9	150	50
10	200	50
11	250	50
12	300	100
13	400	100
14	500	100
15	600	100
16	700	100
17	800	100
18	900	100
19	1000	100
20	1100	100
21	1200	100
22	1300	100
23	1400	100
24	1500	250
25	1750	250
26	2000	250

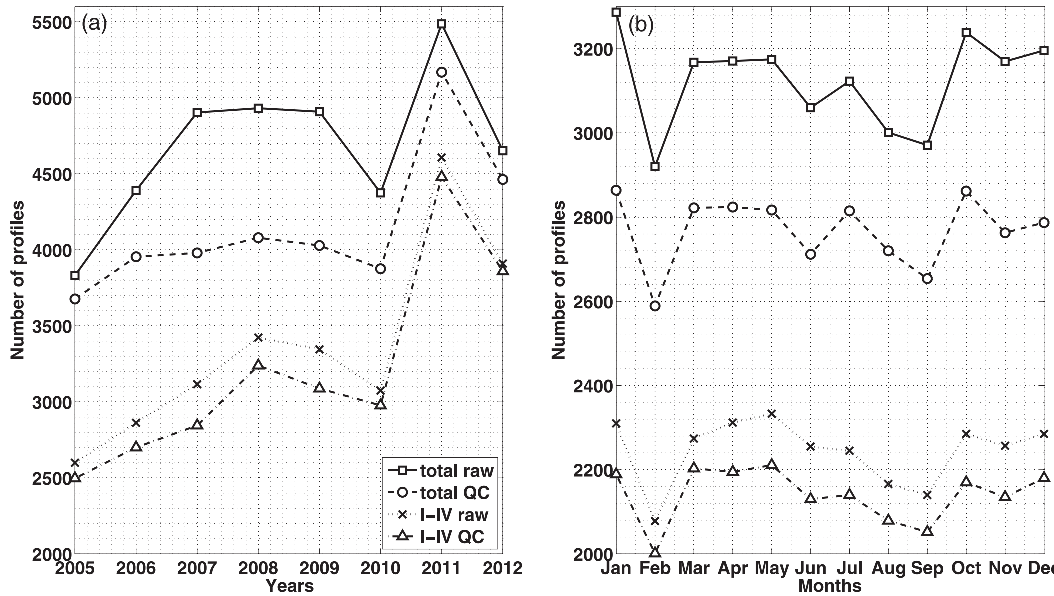

FIG. 2. Number of Argo profiles by (a) year and (b) month during the period 2005–12 over the entire study domain and the Kuroshio region (subregions I–IV in Fig. 1). Solid line with squares (dotted line with × symbols) denotes the number of raw profiles in the total region (subregions I–IV). Dashed line with circles (dashed line with triangles) indicates the number of profiles after the final QC in the total region (subregions I–IV).

representativeness errors (Böhme and Send 2005; Johnson et al. 2013).

The OA technique by Carter and Robinson (1987), which is based on the Gauss–Markov theorem, is chosen in this study to compute the gridded climatology while minimizing the variance of a random variable at each grid (i.e., reducing the random noise influence). We use the Gaussian correlation function, assuming the targeted fields are stationary and homogenous to reasonably simplify the OA algorithm (Montanari et al. 2006). Given the data coverage, the OA parameters (i.e., grid size of ¼°, decorrelation scale of 200 km, and *e*-folding decay scale of 120 km) are determined from a series of OA sensitivity experiments, by quantifying the OA mapping error (i.e., root-mean-square differences between the mapped variable and the sampled variable). Meanwhile, a smoothing filter (e.g., a five-point Laplacian filter) is applied in the OA fields. In previous studies the OA technique has been applied to Argo data at a coarser resolution, for example, 3° in the Pacific as a whole (Zhang et al. 2013) and ½° in the Japan Sea alone (Park and Kim 2013). To the best of our knowledge, this is the first application of OA to Argo data over the complex WBC region at a horizontal resolution as high as ¼°. This allows us to study not only the general oceanic circulation but also mesoscale features that play an important role in oceanic mixing and transport of heat, water, momentum, and nutrients, etc. (e.g., Zhang et al. 2001; Fu et al. 2010; Qiu and Chen 2010; Chang and Oey 2011).

3. Results

a. QC-processed profiles

Figure 2 summarizes the number of Argo profiles before and after the final QC method in each year (Fig. 2a) and each month of the period 2005–12 (Fig. 2b). In the entire study region, ~89% of raw profiles passed the final QC (i.e., the final QC including both individual and climatological checks), whereas in the subregions I–IV, the rate grew up to ~95%. In subregion V (i.e., part of the Japan Sea), only ~72% of the original profiles passed the final checks, mainly due to the pressure errors from the wide deployment of faulty pressure sensors before 2010 (Barker et al. 2011), which can be identified from the information in the Argo files. Figure 2a shows an overall increase in the quantity of good profiles during 2005–12 in the whole study region, except for a sudden drop in 2010 due to the problematic APEX sensor halt and recall in 2009 (Argo Steering Team 2011; Barker et al. 2011). Since 2010 fewer profiles have been filtered out by the QC method (Fig. 2a), consistent with technical advancements: for example, the replacement of faulty sensors and the yearly expansion of Iridium technology employed in the floats to reduce data transmission errors (Argo Steering Team 2011). Figure 2b

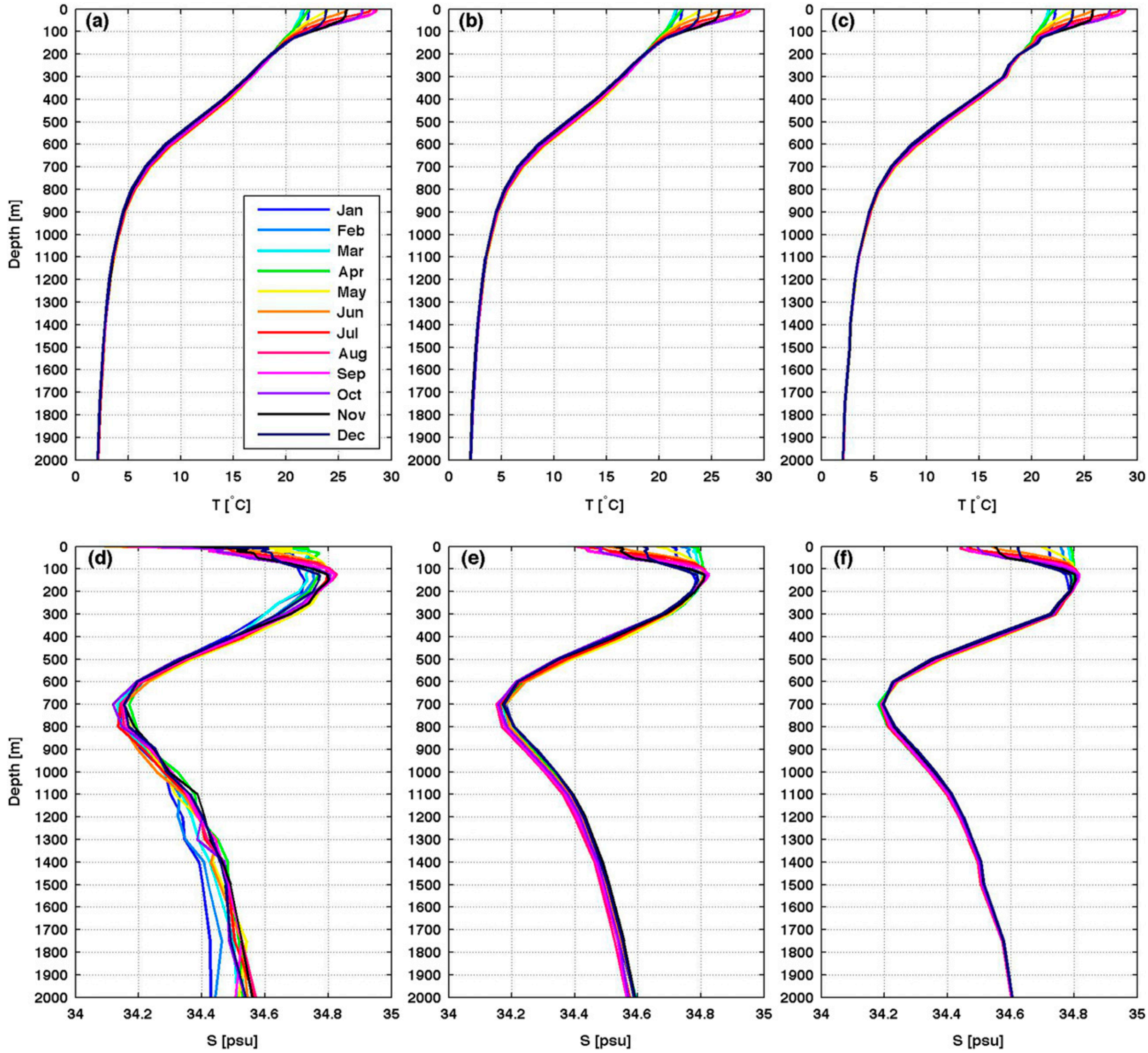

FIG. 3. Area-averaged monthly mean Argo profiles (2005–12) for (a)–(c) temperature and (d)–(f) salinity over the Kuroshio region (subregions I–IV in Fig. 1). (a),(d) Raw data; (b),(e) data after individual QC only; and (c),(f) data after the final QC.

exhibits that the Argo data quality in both cold and warm seasons was improved by our method (by the effective removal of nonrepresentative data in each monthly climatology) over the Kuroshio region and the whole study domain, where atmospheric conditions (e.g., extratropical storms, tropical cyclones, and typhoons) may play a potential role in challenging the Argo observation design and maintenance in the complex western boundary region.

The monthly climatologies of area-averaged hydrographic profiles over the Kuroshio region before and after QC are presented in Fig. 3. The processed profiles (after the final QC) shown in Fig. 3c for temperature and Fig. 3f for salinity are more physically reasonable than the raw profiles (Figs. 3a and 3d) and the profiles merely after the individual QC (Figs. 3b and 3e). In particular, the climatological check (i.e., collective QC) improves the quality of salinity profiles at the deep depths below the Kuroshio Thermocline and Intermediate Waters, where by nature very little seasonal variation should be experienced. The salinity errors, as shown in Fig. 3d for the raw monthly mean profiles, are greater at deeper depths (e.g., below 1000 m) than at upper depths (e.g., above 700 m). The relatively large deep-layer errors are primarily due to the vertical shifts of the pressure profiles by the aforementioned faulty pressure sensors

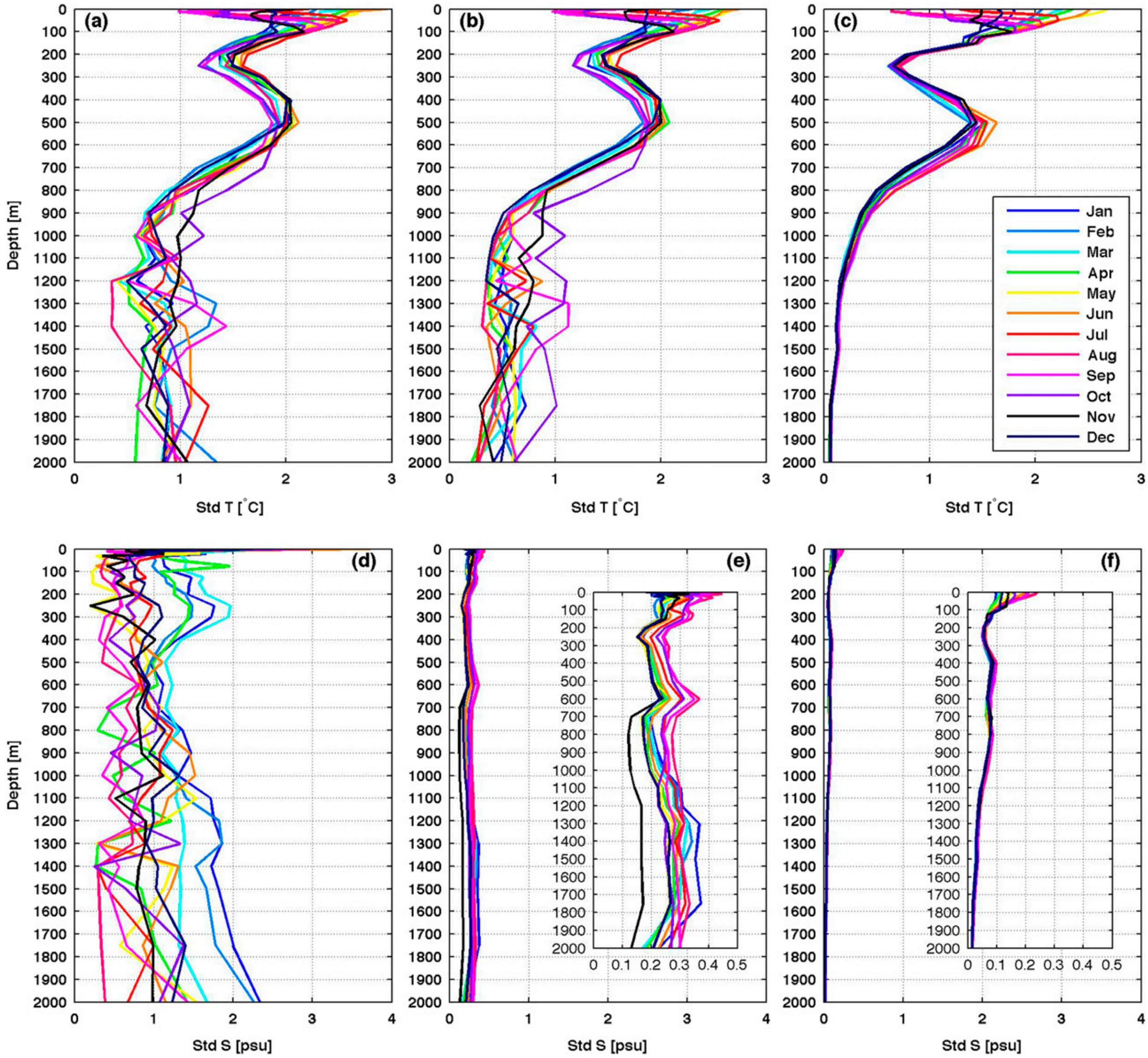

FIG. 4. As in Fig.3, but for the standard deviations of the monthly Argo profiles. Inset in (e),(f) shows the standard deviation of the area-averaged monthly salinity profiles zoomed in from 0 to 0.5 psu for the data after individual QC and after the final QC, respectively.

deployed before 2010. The vertical pressure biases could affect both salinity and temperature observations (Barker et al. 2011). In this study the temperature profiles (e.g., Figs. 3a–c) are in good agreement between originally observed profiles and quality-controlled profiles, because temperature is not as much spatially varied as salinity in the western boundary areas and temperature has fewer observational problems in techniques than salinity (e.g., salinity drifts or offsets from biofouling). However, the temperature errors due to the vertical pressure shifts indeed exist and are more evident in the standard deviation of the corresponding raw profiles (Fig. 4a). It is necessary to carry out the QC procedures for all the hydrographic profiles before any WBC research, and the monthly climatology of each subregional profile is effectively improved in data quality by our method. The standard deviations of the monthly mean profiles for temperature and salinity are displayed in Fig. 4, corresponding to Fig. 3, respectively, for raw data (Figs. 4a,d), individual QC (Figs. 4b,e), and the final QC (Figs. 4c,f). It is well known that salinity is a better tracer than temperature in the WBC region, where there exist intrinsically large salinity variations. But it poses more difficulties for measuring and acquires greater errors than those from temperature as suggested in Figs. 3d and 4d in comparison to Figs. 3a and 4a. The large standard

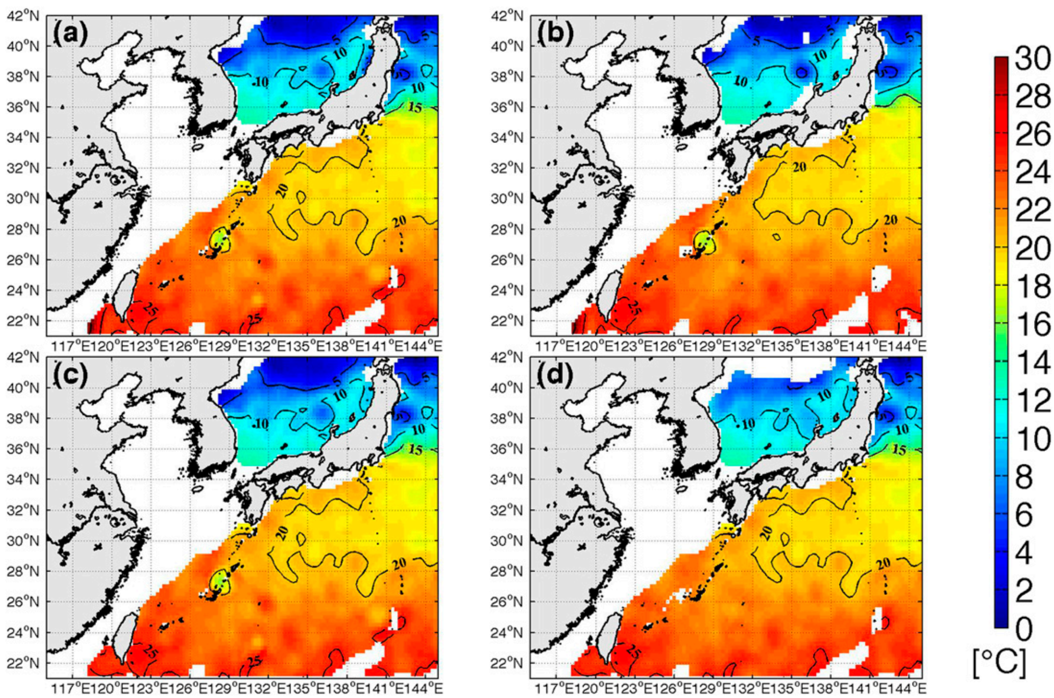

FIG. 5. The OA distribution of sea surface temperature in February (2005–12) from Argo profiling data with four check methods: (a) individual QC only, (b) individual QC and collective QC (climatological check) of regular 1° × 1° bins, (c) individual QC and collective QC of the entire region, and (d) individual QC and collective QC of subregions.

deviations in the deep layers observed in the raw and individual QC profiles for both temperature and salinity are greatly reduced by the final QC, suggesting that our method has value both in defining a climatological mean and in identifying outlier data.

b. OA-gridded fields

Using the same OA technique (in section 2b) of mapping three-dimensional distributions of temperature and salinity for each month, we perform a series of sensitivity experiments with different QC methods to check the spatial continuity or consistency of adjacent points from the monthly climatological Argo datasets. The appropriate QC procedure is determined based on an analysis of the results of different QC methods.

Figure 5 is an example of temperature distribution at the sea surface in February, defined as the upper-5-m bin-averaged depth measurements (Table 1). The climatological check can be conducted with either regular grids or irregular areas (Manzella and Gambetta 2013). Thus, we could find a suitable method for the monthly climatological check through the sensitivity test for different approaches: 1) in each 1°×1° (or 2° × 2°) regular grid, 2) in the entire region, and 3) in each of the irregular subdivisions (see Fig. 1; section 2a). By comparing with the sea surface temperature field in February after the individual QC (Fig. 5a), the application of a climatological check with regular grid bins (e.g., 1°× 1° in Fig. 5b) tends to be too rigorous to overcontrol more ''bad'' data, which are probably ''good,'' and leads to even worse distribution with severe bull's-eyes found in many places in the study region. Whereas a climatological check over the entire region seems to undercontrol the data, seeing as there is very little change in the field in comparison with Figs. 5c and 5a. The climatological check of the irregular subregions with different sizes and geometries gives the smoothest field (Fig. 5d) in comparison to the above-mentioned approaches (Figs. 5a–c). The check focuses on specific coherent subregions, to implement QC in a manner such that more details about the subregional dynamics are taken into account. In particular, this method is capable of removing the singularity along the Ryukyu Islands, for instance. The method (in Fig. 5d) could give significantly improved Argo data distributions in the Kuroshio region, though the upper-5-m Argo data in this region were usually regarded as not good as to be used in previous studies (Zhang et al. 2013). Similar findings hold for the distribution of sea surface salinity (Figs. 6a–d) in the warm season (e.g., August), when the data quality could have more problems, since greater variations of salinity are distributed over the Kuroshio region in comparison

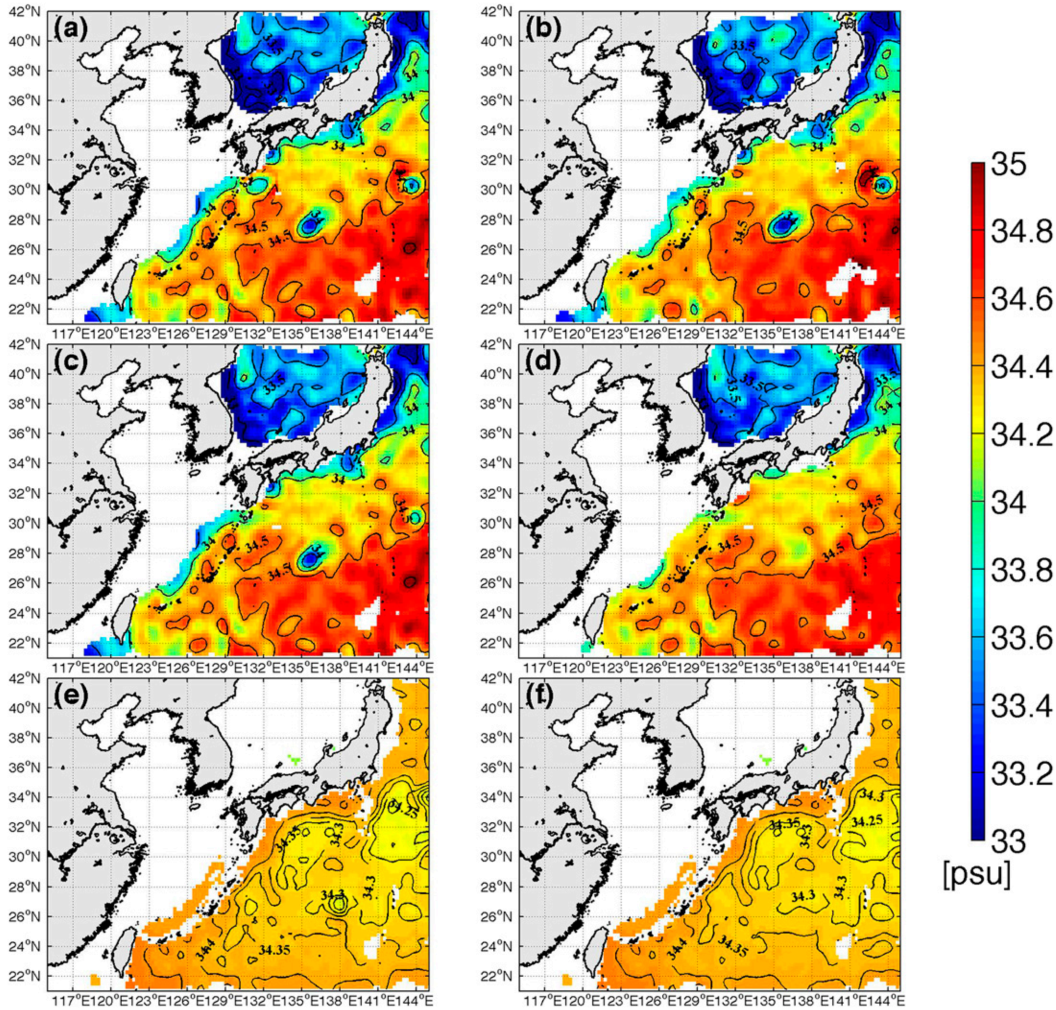

FIG. 6. (a)–(d) As in Figs. 5a–d, respectively, but for sea surface salinity in August. (e) Salinity distributions at 1000-m depth with the same QC method as in (d), but applied to the collective QC only once; (f) as in (e), but applied to the climatological check twice.

to other seasons. This demonstrates that our method also works for the more difficult case (e.g., salinity in August) in the study region.

Then we check further the number of iterations needed for the climatological check. To show the effects, we display the salinity distributions at 1000 m (Figs. 6e and 6f) where the successive climatological check iterations are applied once (Fig. 6e) or twice (Fig. 6f). As well as the upper depths, we also investigate the deeper depths, since evident sampling errors (e.g., bull's-eyes) tend to appear more frequently at the upper depths, where there usually exist much larger physical variations than at the deeper depths. The deep-layer errors (e.g., around 26°–28°N, ~138°E in Fig. 6e), most likely correlated with the upper-layer bull's-eyes (e.g., around 26°–29°N, 134°–138°E in Figs. 6a–c), may exist randomly in space and time. These errors could be aroused from the bottom topographic impacts and the aforementioned faulty pressure sensors. A comparison of Figs. 6e and 6f suggests that the second-time application of the climatological check more effectively removes structures in the field that can be considered noises. Ideally, it is necessary to demonstrate convergence of standard deviation by making more iterations. In practice, however, this is computationally prohibitive so far. And to strike a balance between keeping sufficient data and carrying out adequate quality controls, we applied the climatological check twice. Nevertheless, the results after two iterations seem physically reasonable (from Fig. 6f and other spatial analyses for all 12-month climatologies), computationally adequate, and strict enough (e.g., see Fig. 7; more data are flagged by narrowing down the standard deviation in comparison to other methods). The two time iterations, also widely used by many other studies

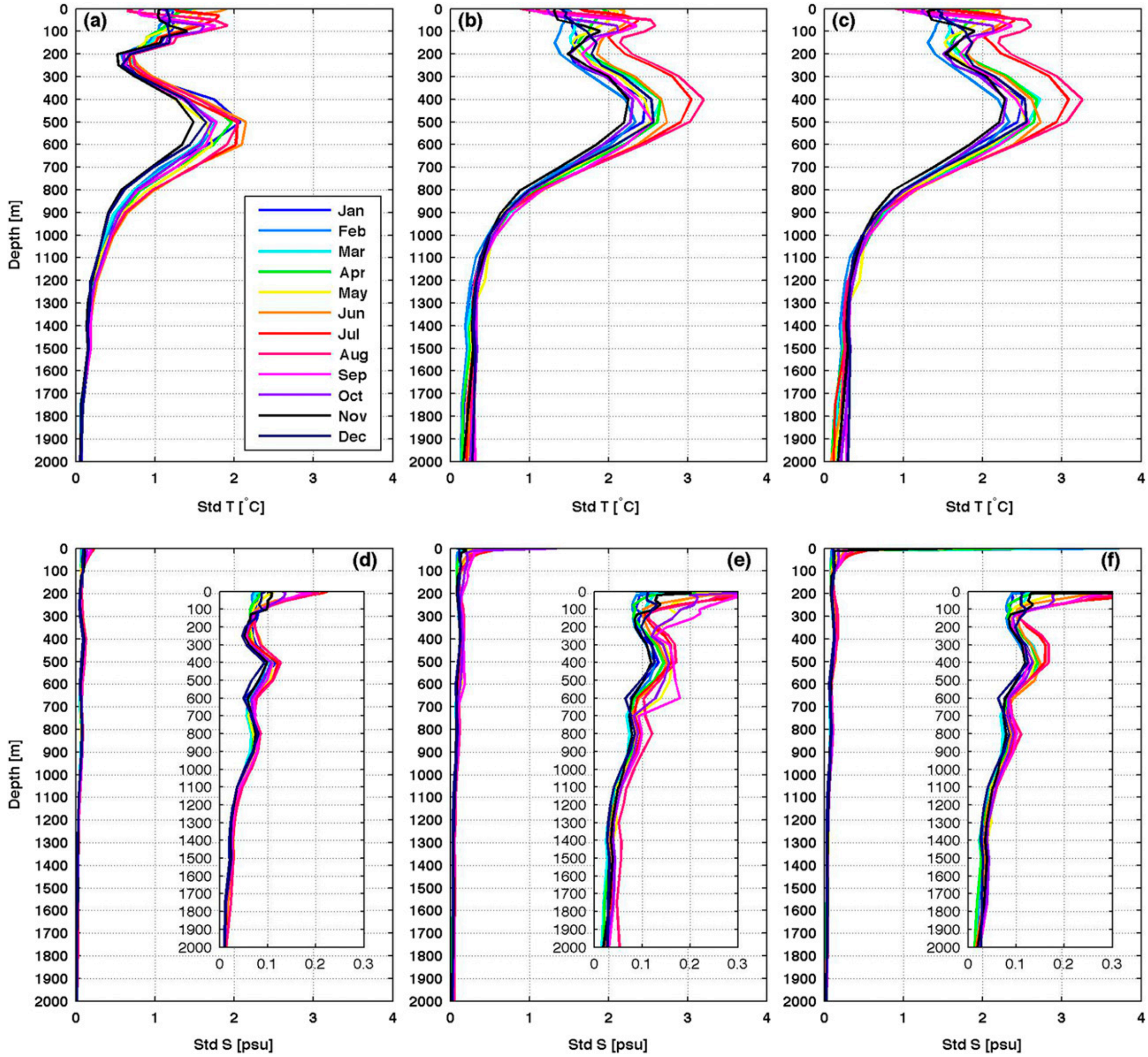

FIG. 7. Comparisons of the standard deviation of monthly averaged Argo profiles (2005–12) over subregion IV (Fig. 1) between the final QC for (a) temperature and (d) salinity and two other QC Argo datasets: WOD13 [(b) temperature; (e) salinity] and GDACs [(c) temperature; (f) salinity]. Inserts in (d)–(f) are standard deviations of the area-averaged monthly salinity profiles zoomed in from 0 to 0.3 psu for the data from the final QC, the WOD13, and the GDACs, respectively.

(e.g., Montanari et al. 2006; Johnson et al. 2013; Locarnini et al. 2013; Zweng et al. 2013), can efficiently eliminate random errors and nonrepresentative data.

4. Evaluations and discussion

a. *Evaluation of profiling data with other QC profiling datasets*

The final QC monthly mean vertical profiles are compared with the corresponding Argo profiles processed by two other QC approaches: 1) the WOD13 profiles flagged 0 (''accepted'' value; Boyer et al. 2013) and 2) the GDAC profiles flagged 1 (''good'' value; Argo Data Management Team 2012).

Figure 7 shows the standard deviation of the monthly climatological profiles averaged over the Kuroshio subregion IV from our final QC, WOD13, and GDACs. Our QC method shows smaller standard deviations at almost every depth for both temperature (Fig. 7a) and salinity (Fig. 7d) with respect to WOD13 (Figs. 7b and 7e) and the GDACs (Figs. 7c and 7f). Performing a QC on specific coherent subregions, our QC method checks

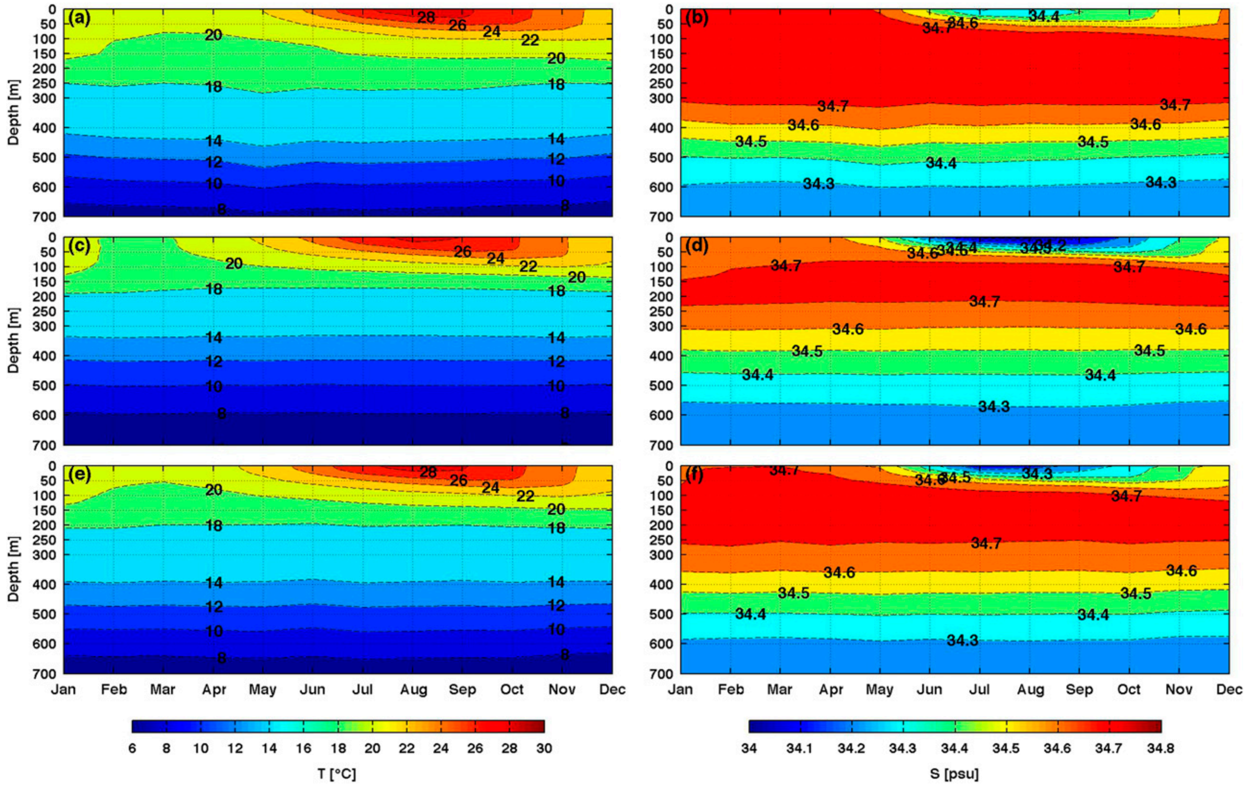

FIG. 8. Evaluation of the OA climatology (2005–12) area-averaged seasonal evolution of (a) vertical temperature and (b) salinity, with those from two coordinated databases: RA [(c) temperature; (d) salinity] and *WOA13* [(e) temperature; (f) salinity], over subregion IV.

the Argo data more efficiently in consideration of more detailed local hydrodynamics, whereas the other two methods are not focusing on our study region and apply the climatological check at a relatively loose-measured level—for example, at a larger regular grid or a larger threshold for the standard deviation. Similar results are found for other subregions (not shown). Among all the subregions, in subregion IV our method (Fig. 7a) shows the most distinct standard deviation for temperature in comparison with the other two methods (Figs. 7b and 7c). This indicates that our QC method designed specifically for the Kuroshio region may delineate some new regional hydrographic and hydrodynamic features (e.g., for the thermocline layers above ~700 m; Fig. 7a) in comparison to the traditional QC from WOD13 or the GDACs. For salinity, the standard deviation of the GDACs' flagged profiles tends to be greater at the sea surface (Fig. 7f) than those from our method (Fig. 7d) and WOD13 (Fig. 7e). Overall, the QC-processed profiles from the three methods are within similar ranges, with comparable monthly averaged profiling structures in the near-surface, subsurface, and intermediate layers. It is thus reasonable to use any of the QC methods, and the detailed procedures could be varied depending on the purpose of the specific study.

b. Assessing climatological fields with other coordinated datasets

The gridded Argo data after the final QC method, called OA climatology, are then assessed with the *WOA13* and RA, in terms of both vertical and horizontal hydrographic distributions on the same spatial (1/4°) and temporal (2005–12) scales.

Figures 8 and 9 show the seasonal evolution of the upper-layer vertical thermohaline characteristics averaged over the core subregion IV and the Kuroshio region (subregions I–IV) from the three datasets: OA climatology, RA, and *WOA13*. OA climatology has more similarities with *WOA13* than with RA in all the regions, primarily due to the Argo observation (the only data source for OA climatology) accounting for a greater proportion for *WOA13* (a multitype in situ database) than for RA (assimilated satellite remote sensing is another major observation source). Despite the similar seasonal evolution of stratifications (e.g., thermocline, halocline, and implied pycnocline, particularly in the warm season in Figs. 8 and 9) in the

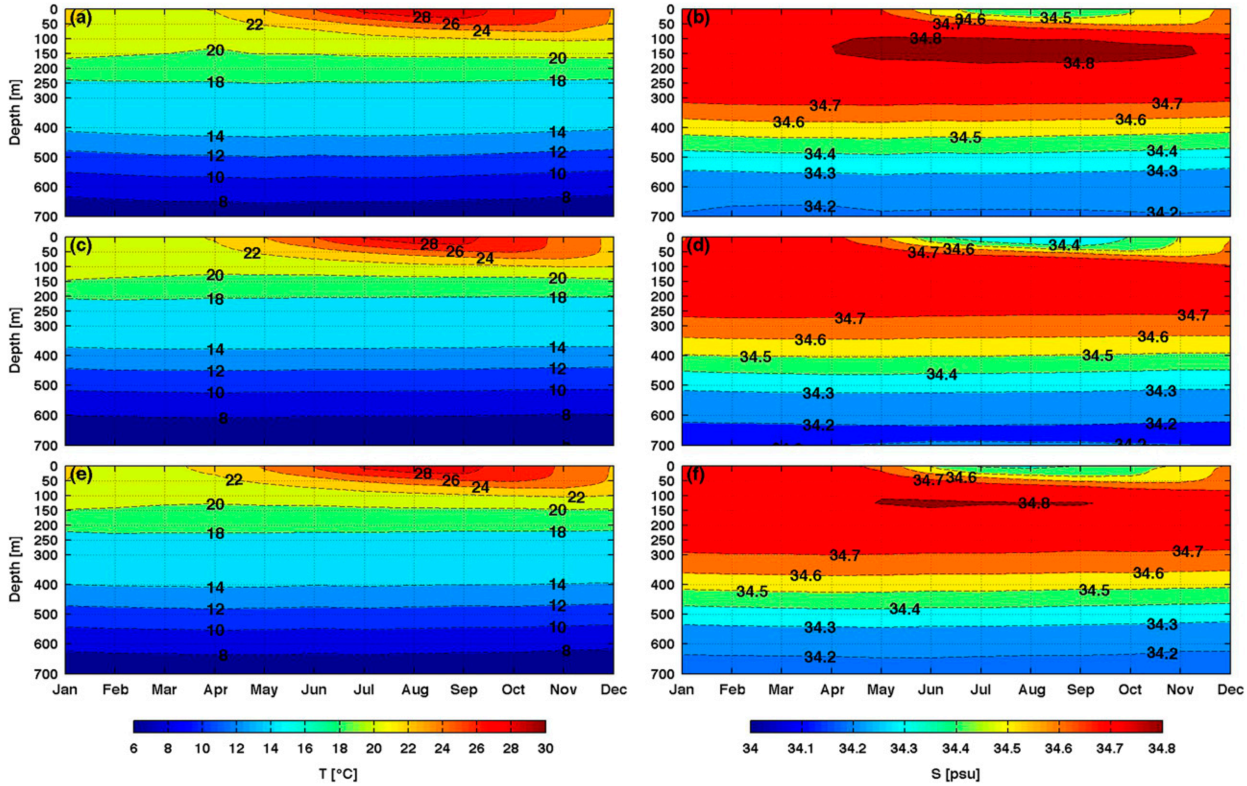

FIG. 9. As in Fig. 8, but for the Kuroshio region (subregions I–IV).

Kuroshio region (and in subregion V the part of the Japan Sea, not shown), OA climatology (Figs. 8a and 8b) shows stronger ventilation and more expansive surface mixed layer (by Ekman pumping via atmospheric forcings) in the cold season than RA (Figs. 8c and 8d) and *WOA13* (Figs. 8e and 8f). This comparison suggests that the mixed layer tends to be flagged by the traditional QC (from GDACs, *WOD13*, etc., that contribute to the coordinated databases). Hence, OA climatology is expected to provide more details about the mixed layer and ventilated processes in the WBC region. OA climatology also shows slightly warmer and saltier Kuroshio regional averaged seawater from near-surface down to intermediate layers in nearly every month than RA and *WOA13* (Figs. 8 and 9). The subsurface saltier band (e.g., 34.8 psu in Fig. 9) in the warm season, as a result of area-averaged STCC and related mode waters over the whole Kuroshio region (subregions I–IV), could be observed more conspicuously from OA climatology (Fig. 9b) than from RA (Fig. 9d) and *WOA13* (Fig. 9f). It implies that our OA climatology (with the QC designed specifically for the WBC region) may capture more hydrographic and hydrodynamic details than the traditional coordinated databases, *WOA13* and RA. And the distinct salinity behavior with more details observed from OA climatology may play a critical role in studying the water cycle and related climatic issues over the complex WBC region. Although combined multitype measurements (including more data to describe fields in more detail and with better precision) are expected to generate better gridded fields than those single-type observing systems (e.g., Schmid 2005), they would also introduce time disparity and other issues between various types of observing systems.

The warm-season (e.g., August) horizontal salinity distributions of OA climatology is evaluated in terms of the differences between any two of the three datasets: OA climatology, RA, and *WOA13*. Figure 10 shows the comparisons at selected depths (surface, 50 m, 100 m, and 200 m) of the upper ocean, where major currents (Kuroshio, STCC, etc.) and other energetic phenomena (e.g., mesoscale eddies) are primarily located, with much stronger atmospheric influences than at deeper depths. OA climatology from the Argo observing system alone is similar to both RA and *WOA13*, especially for the areas (in light colors and close to zero value in Fig. 10) with abundant observational points (Fig. 11), as Argo is the dominant observation in the coordinated datasets. The bigger

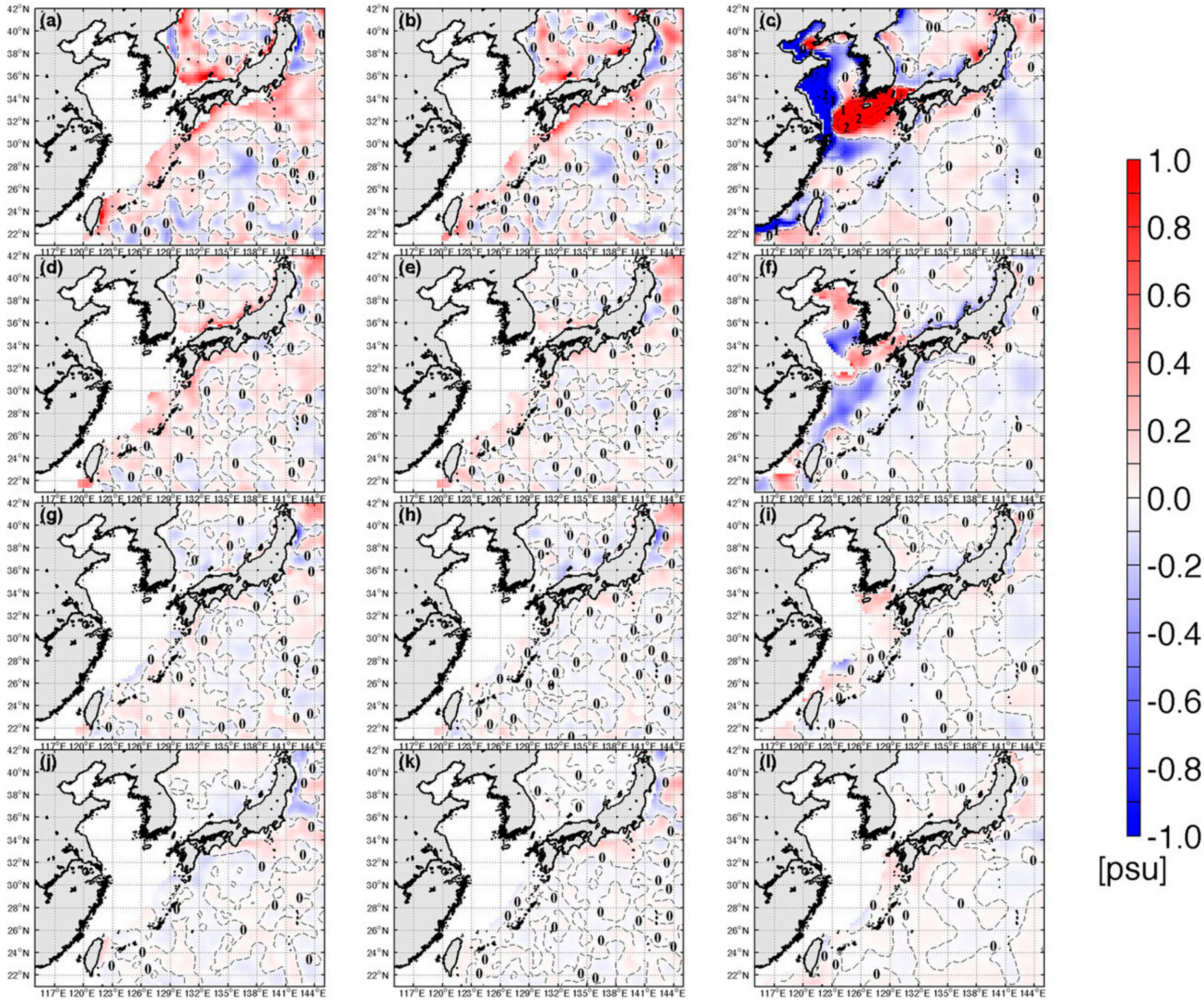

FIG. 10. Differences of the salinity distributions in August (2005–12): (a),(d),(g),(j) between OA climatology and RA; (b),(e),(h),(k) between OA climatology and *WOA13*; and (c),(f),(i),(l) between RA and *WOA13*; at the (a)–(c) surface, (d)–(f) 50-, (g)–(i)100-, and (j)–(l) 200-m depths.

differences mainly occur at the upper-50-m depths (dark colors in Figs. 10a–f) near the coasts and boundaries (islands), etc., where traditional CTD measurements are twice those of the Argo profiles (Fig. 11). The number of Argo observations in OA climatology accounts for only ~40% of the number of the whole in situ observations in *WOA13*. The largest differences occur between RA with *WOA13* (Figs. 10c and 10f) in the marginal China seas (e.g., East China, Yellow, and Bohai Seas) and the straits (e.g., Tsushima and Taiwan Straits), where available observations are scarce. Because of the insufficient number of data (e.g., Fig. 11), mapping of the hydrographic fields (especially for salinity; Fig. 10) still produces quite uncertain results for the marginal China seas and the nearby straits, even if that mapping is based on combined in situ datasets (e.g., *WOA13*) or assimilated in situ and satellite data (e.g., RA).

Tables 2 and 3 summarize the subregional statistics from the three datasets (OA climatology, RA, and *WOA13*) for temperature and salinity at the sea surface, where there are usually the largest differences between the different datasets (e.g., Fig. 10) in both cold (February) and warm (August) seasons. Consistent with the foregoing vertical analyses of temperature and salinity, warmer and saltier water are observed from OA climatology for both seasons in the core subregion IV, where the spatially averaged differences (DIFF) are slightly positive. The agreement between OA climatology and the coordinated datasets from the subregional statistics is reasonably good—for example, in terms of spatial distributions—with a correlation (CORR) of above 0.6 (statistically significant at

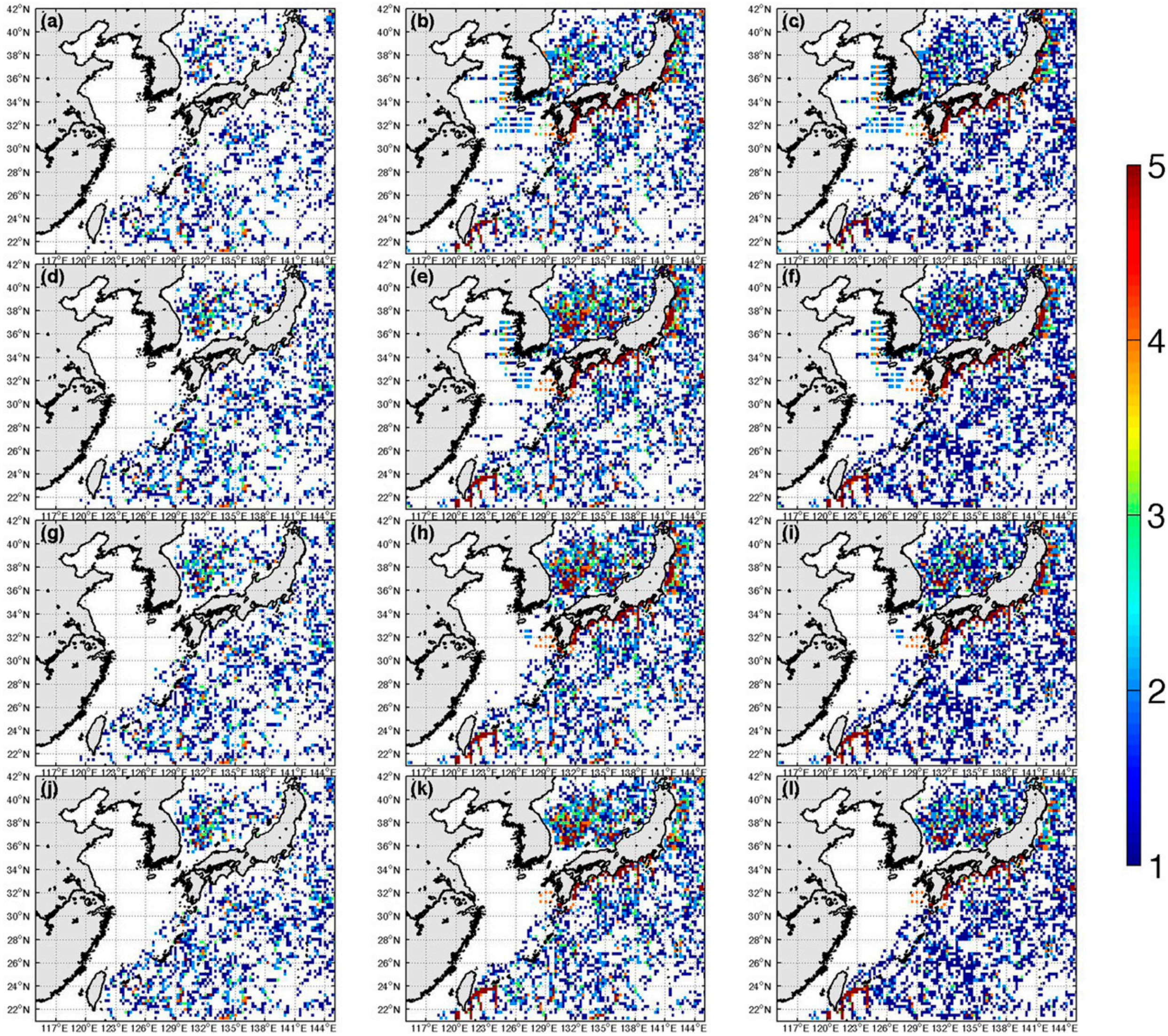

FIG. 11. Comparisons of the number of observations in each 1/4° grid box among Argo data in (a),(d),(g),(j) OA climatology; (b),(e),(h),(k) all the in situ measurements used in *WOA13*; and (c),(f),(i),(l) all non-Argo measurements used in *WOA13* for salinity in August (2005–12); at the (a)–(c) surface, (d)–(f) 50-, (g)–(i) 100-, and (j)–(l) 200-m depths.

the 95% confidence level) for most subregions. The low correlation in some subregions with relatively small area (see Fig. 1; e.g., subregion I or II) is primarily due to less data availability for statistics calculations, though the data quality matters, too. Thus N, the number of data points available for calculation, is also shown in the tables. In comparing RMSE2 with RMSE, the magnitude of systematic error of OA climatology relative to RA (or *WOA13*) can be inferred as negligibly small in all the subregions. The differences between OA climatology and the coordinated datasets in statistics are mainly due to the inconsistency of observational data from the different datasets. Hence, it is expected that enhancement of Argo observations could further improve the agreement between OA climatology and the coordinated datasets.

5. Conclusions and implications

A comprehensive QC procedure has been developed to process multiyear (2005–12) Argo profiling data over the energetic WBC (Kuroshio) region, one of the most important areas in the World Ocean. The method includes both individual and climatological (i.e., collective) checks, along with OA technique, to generate gridded climatologies. The collective QC is applied to specific coherent subregions; hence, more details about the local hydrodynamics are taken into

TABLE 2. Summary of the five subregions (I–V) statistics for comparison of sea surface temperature between OA climatology and RA (or *WOA13*) in winter (February) and summer (August). DIFF is the spatially averaged differences between OA climatology and RA (or *WOA13*). CORR is the spatial correlation coefficient of OA climatology and RA (or *WOA13*), an indicator of similarity in the geographical distribution between two datasets, and all the correlation coefficients are statistically significant at the 95% confidence level. RMSE is the root-mean-square error of OA climatology relative to RA (or *WOA13*), quantifying the absolute differences between two datasets. RMSE2[a] is the root-mean-square error excluding the systematic error of OA climatology relative to RA (or *WOA13*).

	Temperature		DIFF	CORR	RMSE	RMSE2	*N*
I	OA–RA	Feb	0.015	0.8300	0.5973	0.5971	733
		Aug	0.001	0.2044	0.2903	0.2903	723
	OA–*WOA13*	Feb	−0.043	0.8850	0.4920	0.4901	733
		Aug	0.094	0.5009	0.2873	0.2716	723
II	OA–RA	Feb	0.027	0.6874	0.6365	0.6359	736
		Aug	−0.143	0.1579	0.4172	0.3921	755
	OA–*WOA13*	Feb	−0.026	0.7619	0.5510	0.5504	736
		Aug	−0.087	0.4586	0.3580	0.3473	755
III	OA–RA	Feb	0.089	0.9113	0.5761	0.5692	1036
		Aug	0.135	0.5567	0.3306	0.3016	1033
	OA–*WOA13*	Feb	−0.008	0.9420	0.4662	0.4661	1036
		Aug	0.033	0.6779	0.2475	0.2454	1033
IV	OA–RA	Feb	0.285	0.8747	0.7643	0.7091	1444
		Aug	0.098	0.6842	0.4063	0.3942	1442
	OA–*WOA13*	Feb	0.092	0.8868	0.7202	0.7143	1444
		Aug	0.093	0.7057	0.3744	0.3626	1442
V	OA–RA	Feb	−0.007	0.9039	1.0163	1.0162	779
		Aug	0.004	0.8125	0.8688	0.8687	913
	OA–*WOA13*	Feb	0.112	0.9132	0.9390	0.9323	779
		Aug	0.271	0.8056	1.0158	0.9790	913

[a] RMSE2 = $\sqrt{(1/N)\sum_{i=1}^{N}[(x_i - y_i) - (\overline{x} - \overline{y})]^2}$, where x and y denote the two datasets OA climatology and RA (or *WOA13*). The N is the number of data points used for the statistics calculations.

account. In particular, climatological values and related standard deviations used in the collective QC method are shown to be better if defined in regime-oriented subregions I–V instead of a large-scale area or in a spatially uniform grid. By eliminating the nonrepresentative data and random noises more efficiently, both processed profiles and resultant climatological fields are improved effectively. Our method has

TABLE 3. As in Table 2, but for salinity.

	Salinity		DIFF	CORR	RMSE	RMSE2	N
I	OA–RA	Feb	−0.015	0.6587	0.0774	0.0759	652
		Aug	0.026	0.3938	0.2112	0.2095	668
	OA–*WOA13*	Feb	0.009	0.7702	0.0708	0.0702	652
		Aug	0.039	0.6450	0.1420	0.1364	668
II	OA–RA	Feb	0.019	0.6352	0.0750	0.0726	730
		Aug	0.001	0.2977	0.1089	0.1089	745
	OA–*WOA13*	Feb	0.015	0.7125	0.0688	0.0671	730
		Aug	0.010	0.5251	0.0993	0.0988	745
III	OA–RA	Feb	−0.001	0.4373	0.0642	0.0642	1037
		Aug	0.012	0.7211	0.1294	0.1289	1038
	OA–*WOA13*	Feb	−0.003	0.6879	0.0516	0.0515	1037
		Aug	−0.022	0.8595	0.0947	0.0921	1038
IV	OA–RA	Feb	0.010	0.7467	0.0532	0.0523	1438
		Aug	0.120	0.6951	0.1943	0.1532	1354
	OA–*WOA13*	Feb	0.040	0.6259	0.0891	0.0798	1438
		Aug	0.086	0.7377	0.1850	0.1640	1354
V	OA–RA	Feb	0.042	0.7046	0.0845	0.0734	946
		Aug	0.130	0.5949	0.2942	0.2639	918
	OA–*WOA13*	Feb	0.085	0.6360	0.1696	0.1468	946
		Aug	0.114	0.5974	0.2590	0.2326	918

value both in defining a climatological mean and in identifying outlier data.

In an assessment with other QC approaches (from GDACs and WOD13) and coordinated datasets (RA and *WOA13*), the agreement is reasonably good in terms of QC-processed profiling data, and vertical and horizontal distributions, especially for those areas with abundant observations. More detailed or mesoscale features in the horizontal and vertical fields (e.g., stronger ventilation and a larger expansion of the mixed layer) may be inferred by our QC method, which was designed specifically for the study region, than by the traditional QC-involved, more smoothly distributed coordinated datasets, *WOA13* or RA. Though combined multitype measurements are expected to generate better gridded fields than those single-type observing system (Schmid 2005), they would also introduce time disparity and other issues between various types of observing systems. The disagreement between OA climatology and the coordinated datasets are likely due to the inconsistency of observational data from the different datasets, and a higher degree of agreement hence would be expected after enhancement of Argo observations. Presently there still exist great uncertainties in describing climatological fields in the marginal China seas and the nearby straits, even from combined in situ datasets (e.g., *WOA13*) or assimilated in situ and satellite data (e.g., RA), based on the very limited observations there. The possible future expanding in situ observations (particularly Argo) into these interior marginal seas would be expected to significantly improve mapping the three-dimensional fields, to better address WBC-connected dynamical and climatic issues.

The QC could be applied to other datasets and other regions of the global ocean where specific coherent subregions of dynamical importance should then be defined to assess the appropriate standard deviations for quality control. Our method holds the potential to meet future challenges in processing likely growing Argo observations in the coming decades, too. In addition, the processed and analyzed data could also facilitate numerical simulations in several aspects (data assimilation, initial condition, reference climatology, etc.), especially for the WBC region, which is of much scientific importance but is usually not simulated as well as other regions of the World Ocean.

Acknowledgments. The authors were supported by the GEMINA Italian project for the OOCC China–Italy collaboration. WJ and DW were also supported by the Startup Foundation for Introducing Talent of NUIST. We thank three anonymous reviewers for their helpful suggestions. We had helpful comments from P.-M. Poulain, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale. We particularly appreciate the U.S. National Oceanic and Atmospheric Administration for the World Ocean Database and *World Ocean Atlas*. The Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it (http://www.argo.ucsd.edu; http://argo.jcommops.org). The Argo program is part of the Global Ocean Observing System.

REFERENCES

Argo Data Management Team, 2012: Argo quality control manual, version 2.7. Argo Data Management, 47 pp. [Available online at http://www.argodatamgt.org/content/download/341/2650/file/argo-quality-control-manual-V2.7.pdf.]

Argo Science Team, 2012: Fourth Argo Science Workshop Meeting and final round table summary. UCSD Rep., 9 pp. [Available online at http://www.argo.ucsd.edu/ASW4_report.pdf.]

Argo Steering Team, 2011: 12th Meeting of the International Argo Steering Team, Buenos Aires, Argentina. UCSD Rep. 120 pp. [Available online at http://www.argo.ucsd.edu/iast12.pdf.]

——, 2013: 14th Meeting of the International Argo Steering Team, Wellington, New Zealand. UCSD Rep., 121 pp. [Available online at http://www.argo.ucsd.edu/iast14.pdf.]

Barker, P. M., J. R. Dunn, C. M. Domingues, and S. E. Wijffels, 2011: Pressure sensor drifts in Argo and their impacts. *J. Atmos. Oceanic Technol.*, **28**, 1036–1049, doi:10.1175/2011JTECHO831.1.

Böhme, L., and U. Send, 2005: Objective analyses of hydrographic data for referencing profiling float salinities in highly variable environments. *Deep-Sea Res. II*, **52**, 651–664, doi:10.1016/j.dsr2.2004.12.014.

Boyer, T. P., and Coauthors, 2009: World Ocean Database 2009. NOAA Atlas NESDIS 66, 216 pp. [Available online at http://www.nodc.noaa.gov/OC5/WOD09/pr_wod09.html.]

——, and Coauthors, 2013: World Ocean Database 2013. NOAA Atlas NESDIS 72, 209 pp. [Available online at http://www.nodc.noaa.gov/OC5/WOD13/.]

Bretherton, F. P., R. E. Davis, and C. B. Fandry, 1976: A technique for objective analysis and design of oceanic experiments applied to MODE-73. *Deep-Sea Res. Oceanogr. Abstr.*, **23**, 559–582, doi:10.1016/0011-7471(76)90001-2.

Carter, E. F., and A. R. Robinson, 1987: Analysis model for the estimation of oceanic fields. *J. Atmos. Oceanic Technol.*, **4**, 49–74, doi:10.1175/1520-0426(1987)004<0049:AMFTEO>2.0.CO;2.

Chang, Y.-L., and L.-Y. Oey, 2011: Interannual and seasonal variations of Kuroshio transport east of Taiwan inferred from 29 years of tide-gauge data. *Geophys. Res. Lett.*, **38**, L08603, doi:10.1029/2011GL047062.

Fu, L.-L., D. Chelton, P.-Y. Le Traon, and R. Morrow, 2010: Eddy dynamics from satellite altimetry. *Oceanography*, **23**, 14–25, doi:10.5670/oceanog.2010.02.

Gaillard, F., E. Autret, V. Thierry, P. Galaup, C. Coatanoan, and T. Loubrieu, 2009: Quality control of large Argo datasets. *J. Atmos. Oceanic Technol.*, **26**, 337–351, doi:10.1175/2008JTECHO552.1.

Johnson, D. R., T. P. Boyer, H. E. Garcia, R. A. Locarnini, O. K. Baranova, and M. M. Zweng, 2013: World Ocean Database

2013 user's manual. S. Levitus, Ed., A. Mishonov, Tech. Ed., NODC Internal Rep. 22, NOAA Printing Office, 172 pp., doi:10.7289/V5DF6P53.

Kobashi, F., H. Mitsudera, and S.-P. Xie, 2006: Three subtropical fronts in the North Pacific: Observational evidence for mode water-induced subsurface frontogenesis. *J. Geophys. Res.*, **111**, C09033, doi:10.1029/2006JC003479.

Locarnini, R. A., and Coauthors, 2013: *Temperature*. Vol. 1, *World Ocean Atlas 2013*, NOAA Atlas NESDIS 73, 40 pp. [Available online at http://data.nodc.noaa.gov/woa/WOA13/DOC/woa13_vol1.pdf.]

Manzella, G. M. R., and M. Gambetta, 2013: Implementation of real-time quality control procedures by means of probabilistic estimate of sea temperature and its temporal evolution. *J. Atmos. Oceanic Technol.*, **30**, 609–625, doi:10.1175/JTECH-D-11-00218.1.

Montanari, G., A. Rinaldi, N. Pinardi, S. Simoncelli, and L. Giacomelli, 2006: *The Currents of Emilia-Romagna Coastal Strip during the Period 1995-2002*. ARPA, 160 pp.

Oka, E., 2005: Long-term sensor drift found in recovered Argo profiling floats. *J. Oceanogr.*, **61**, 775–781, doi:10.1007/s10872-005-0083-6.

Owens, W. B., and A. P. S. Wong, 2009: An improved calibration method for the drift of the conductivity sensor on autonomous CTD profiling floats by θ–S climatology. *Deep-Sea Res. I*, **56**, 450–457, doi:10.1016/j.dsr.2008.09.008.

Park, J. J., and K. Kim, 2013: Deep currents obtained from Argo float trajectories in the Japan/East Sea. *Deep-Sea Res. II*, **85**, 169–181, doi:10.1016/j.dsr2.2012.07.032.

Qiu, B., and S. Chen, 2010: Interannual variability of the North Pacific subtropical countercurrent and its associated mesoscale eddy field. *J. Phys. Oceanogr.*, **40**, 213–225, doi:10.1175/2009JPO4285.1.

Schmid, C., 2005: Impact of combining temperature profiles from different instruments on an analysis of mixed layer properties. *J. Atmos. Oceanic Technol.*, **22**, 1571–1587, doi:10.1175/JTECH1785.1.

Storto, A., and S. Masina, 2014: The new CMCC global ocean eddy-permitting reanalysis (1979-2012): Improvements with respect to the previous version. GEMINA Project Internal Rep., Centro Euro-Mediterraneo sui Cambiamenti Climatici, 9 pp.

Takahashi, T., and Coauthors, 2002: Global sea–air CO_2 flux based on climatological surface ocean $p$$CO_2$, and seasonal biological and temperature effects. *Deep-Sea Res. II*, **49**, 1601–1622, doi:10.1016/S0967-0645(02)00003-6.

Wong, A. P. S., G. C. Johnson, and W. B. Owens, 2003: Delayed-mode calibration of autonomous CTD profiling float salinity data by θ–S climatology. *J. Atmos. Oceanic Technol.*, **20**, 308–318, doi:10.1175/1520-0426(2003)020<0308:DMCOAC>2.0.CO;2.

Yu, L., and R. A. Weller, 2007: Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005). *Bull. Amer. Meteor. Soc.*, **88**, 527–539, doi:10.1175/BAMS-88-4-527.

Zhang, C., J. Xu, X. Bao, and Z. Wang, 2013: An effective method for improving the accuracy of Argo objective analysis. *Acta Oceanol. Sin.*, **32**, 66–77, doi:10.1007/s13131-013-0333-1.

Zhang, D., T. N. Lee, W. E. Johns, C.-T. Liu, and R. Zantopp, 2001: The Kuroshio east of Taiwan: Modes of variability and relationship to interior ocean mesoscale eddies. *J. Phys. Oceanogr.*, **31**, 1054–1074, doi:10.1175/1520-0485(2001)031<1054:TKEOTM>2.0.CO;2.

Zweng, M. M., and Coauthors, 2013: *Salinity*. Vol. 2, *World Ocean Atlas 2013*, NOAA Atlas NESDIS 74, 39 pp. [Available online at http://www.nodc.noaa.gov/OC5/indprod.html.]