MethodsX 8 (2021) 101337

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

An improved method for quality control of *in situ* data from Argo floats using α convex hulls

R. Venkat Shesu[a], T.V.S. Udaya Bhaskar[a,*], E. Pattabhi Rama Rao[a], M. Ravichandran[b], B. Venkateswara Rao[c]

[a] *Indian National Centre for Ocean Information Services (INCOIS), Ministry of Earth Sciences (MoES), Hyderabad, Telangana 500090, India*
[b] *National Centre for Polar and Ocean Research (NCPOR), Ministry of Earth Sciences (MoES), Headland Sada, Vasco-da-Gama, Goa 403804, India*
[c] *Jawaharlal Nehru Technological University (JNTU), Kukatpally, Hyderabad, Telangana 500085, India*

A B S T R A C T

An improved method for detecting abnormal oceanic *in situ* temperature and salinity (T/S) profiles is developed. This procedure extends previous method developed by Udaya Bhaskar et al. [2017].

This method utilizes World Ocean Atlas 2013 gridded climatology which is on 0.25° x 0.25° resolution to build α convex hulls. These α shapes are then used to categorize good and bad *in situ* T/S data profiles. This extended method classify the entire profiles instead of data for standard depths to avoid any errors introduced by interpolation to standard depths. Like in previous method, an 'n' sided polygon (convex hull) encompassing the T/S profile data is constructed using Jarvis March algorithm and Points In Polygon (PIP) principle is employed to judge the profile as good or bad. Extensive sensitivity experiments were done for arriving at the optimal α value such that false positives and true negatives are minimized. All types of issues associated with the *in situ* oceanographic data are identified and quality flag assigned. Examples of this improved method as applied to few Argo floats are presented.

- The T/S profiles corresponding to region of interest are used to build α convex hulls.
- This extended method can be effectively used for quality control of entire profile and clearly demarcate the profile as good/bad.
- This method has the advantage of treating bulk of oceanographic *in situ* profiles data in a single go which filters out erroneous profile data from the good.

* Corresponding author.
E-mail address: uday@incois.gov.in (T.V.S.U. Bhaskar).

https://doi.org/10.1016/j.mex.2021.101337

ARTICLE INFO
Method name: α convex hulls for detecting spurious in situ profiles
Keywords: α Convex Hulls, Point in polygon, Outliers, *In situ* data, Argo floats, Classification
Article history: Received 21 January 2021; Accepted 31 March 2021; Available online 5 April 2021

Specifications Table

Subject area:	Earth and Planetary Sciences
More specific subject area:	Physical oceanography
Method name:	α convex hulls for detecting spurious *in situ* profiles.
Name and reference of original method:	Convex hull method: udaya bhaskar, T.V.S., Seshu, R.V., Timoty, P. Boyer, Rama Rao, E.P., 2017: Quality control of oceanographic *in situ* data from Argo floats using climatological convex hulls, 4, MethodX, 469 - 479. doi.org/10.1016/j.mex.2017.11.007. Jarvish March algorithm: Jarvis, R. A., 1973: On the identification of the convex hull of a finite set of points in the plane, Information Processing Letters, **2,** 18–21. doi:10.1016/0020-0190(73)90020-3. Point-in-Polygon: Implemented using Ray Casting Algorithm. Jordan Curve Theorem: Jordan, C., 1893: Cours D'Analyse de l' Ecole Polytechnique, Paris, second edition.
Resource availability:	World Ocean Atlas 2013. https://www.nodc.noaa.gov/OC5/WOA13/
Submission:	Direct submission

Method details

Background

Argo floats are autonomous CTD profiling floats that drift freely with the sub-surface ocean currents at prescribed parking depths and map temperature and salinity data from a profiling depth to the sea surface at preset time intervals. While the float rises to the surface of the ocean, it measures profiles of conductivity (C) and temperature (T) versus pressure through the water column. This measured temperature and salinity data are reported to various receiving centers via satellites and the autonomous floats sink back to its preset parking depth and continue their new cycles. The Argo program aims to deploy 3000 such autonomous profiling floats with a target profiling depth of 2000 m to observe temperature and salinity within the upper layers of the global ocean, and currents at the parking depths. Each profiling float is expected to have a mean lifespan of ~ 4 years giving good measurements of temperature and pressure, with issues arising with salinity measurements owing to biofouling and other problems [10]. Since the inception of the Argo program in 1999, the geographic distribution of oceanographic T/S profiles data has become more uniform. Abundant data obtained from these profiling floats is being used in operational ocean models, enhancement to existing climatologies [1] and generation of value added products [8].

The oceanic environment is generally harsh for electronic sensors used for mapping, thus making it inevitable to have some spurious measurements. A trustable dataset of observations requires a quality control (QC) procedure capable of detecting spurious data. While manual QC by human experts in the field of ocean sciences minimizes errors, it is incompetent to handle large datasets and vulnerable to inconsistencies between different experts. Udaya bhaskar et al. [9] originally proposed a method for performing quality control of *in situ* data from Argo floats using climatological convex hulls. This method was developed to detect outliers in the profiles measured by Argo profiling floats. However the method proposed to use a convex hull built to individual standard depth levels, requires the observed profiles to be interpolated. This might introduce some error due to interpolation. Also if the spike in the profile data happens to lie between the standard depth they could go unnoticed. In view of this, here we present further improvements to the originally proposed method by using an α convex hull for the entire profile instead of at standard depths. The two main state variables of the ocean, potential temperature, and salinity, S, are related to each other by definite patterns that represent the mean characteristics of a region (e.g., [3,11]). This relation between temperature and salinity (T/S) is used for detecting the erroneous profiles mapped by the profiling floats by building α shapes of T/S profiles.

From examination of many Argo floats, it is clear that the erroneous data or outlier might occur anywhere in the profile and might not be picked up easily by building the convex hull to respective standard depths. Drift in the salinity can be observed only by comparing long term data sets from a single float over a period of one year. TBTO (tributyltinoxide) is used to protect the conductivity cell from biofouling. However, the cell itself can sometimes get contaminated by TBTO, thus resulting in fresh salinity offsets. Usually the TBTO contamination is flushed out after several profiles, and the salinity measurements will return to being correct [7]. The initial profiles appearing to be fresh compared to later profiles owing to Tri-Butyl Tin Oxide (TBTO) may not be effectively observed in convex hulls built to individual depths. Also, from the experience of handling floats since long, it was observed that temperature and pressure sensors are found to be robust. It is only the salinity sensors mounted on Argo floats which are susceptible to degradation owing to bio-fouling [10]. The conductivity cell is more susceptible to fouling and associated drift because of the possible change in the dimension of the conductivity cell due to fouling.

Building the α shapes of T/S profiles

In computational geometry, α-shape is a family of piecewise linear simple curves in the Euclidean plane associated with the shape of a finite set of points. The concept of alpha shape was first defined by Edelsbrunner et al. [2]. The alpha-shape associated with a set of points is a generalization of the concept of the convex hull, i.e. every convex hull is an alpha-shape but not every alpha shape is a convex hull. Fig. 1 gives an illustration of convex hull, α shape and minimal spanning tree. When α is zero, the shape attains the form of minimal spanning tree and when α is infinity the shape turns out to be a convex hull. One needs to fine tune the α such that the resultant figure is neither convex hull nor minimal spanning tree. With this understanding α is fine tuned and α shape were built based on the observations of (i) Temperature vs. Depth, (ii) Salinity vs. Depth and (iii) Temperature vs. Salinity. Once the α shapes are built they are used for detecting and eliminating anomalous profiles either partly of fully.

The T/S data from World Ocean Atlas 2013 on 0.25° x 0.25° resolution was taken and the α shapes of T/S were built for each month, season and annual scale. An '*n*' sided polygon (α shape) with the least area encompassing all the points is constructed using the Jarvis March algorithm [4]. The α convex hulls are built based on an optimal α value fixed, such that the area encompassed by the convex hull is least and either minimal or no points of climatological T/S points are left out during its construction. The final α value used for building the convex hulls (monthly, seasonal and annual) is arrived after an exhaustive iterative process which is also passed through visual inspections to minimize the points being left out of the polygon. Extensive sensitivity experiments were done for arriving at a α, which will provide equal number of false positives and true negatives. Fig. 2 shows the sensitivity experiments done with α and the resulting points falling within α shape (Fig. 2a) and the area of α shape (Fig. 2b).

From the figure one can observe that as α increases, the number of points falling out/left out decreased. After a certain value of α, there seems to be no changes in the number of points left out and the curve looked more or less flat. Using such sensitivity experiments α values were derived for all the months, seasons and annual climatological data sets. These α values are then used to construct the α convex hulls which were eventually used to detect anomalous T/S profiles obtained from various *in situ* sources. The methods for classification of the good and bad profiles is somewhat similar to what is proposed by Udaya Bhaskar et al. [9] and as described below.

Implementation methodology

The principle of α convex hull and Point In Polygon (PIP) implemented using Ray Casting Algorithm [6] are together used to identify good vs bad T/S profiles. The steps for application of the improved method is as follows:

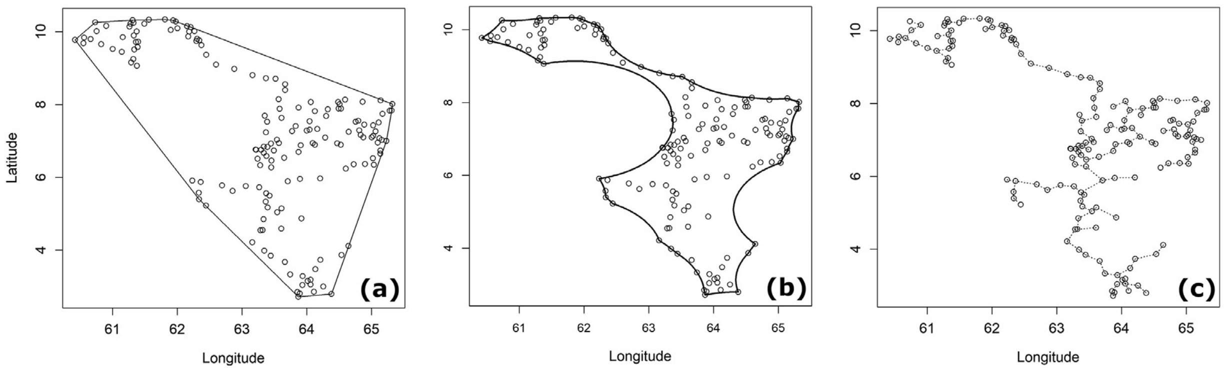

Fig. 1. Example of (a) Convex hull, (b) Alpha shape and (c) minimal spanning tree associated with a set of points pertaining to Argo float trajectory.

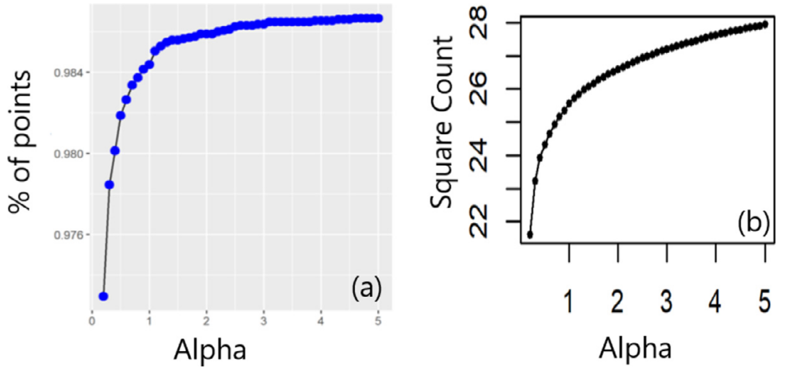

Fig. 2. Sensitivity experiments of α using WOA13 climatology data. (a) Percentage number of T/S data points observed to be falling out of α shape built using T/S, with increasing values of α. (b) Area extent of the α shape with increasing value of α.

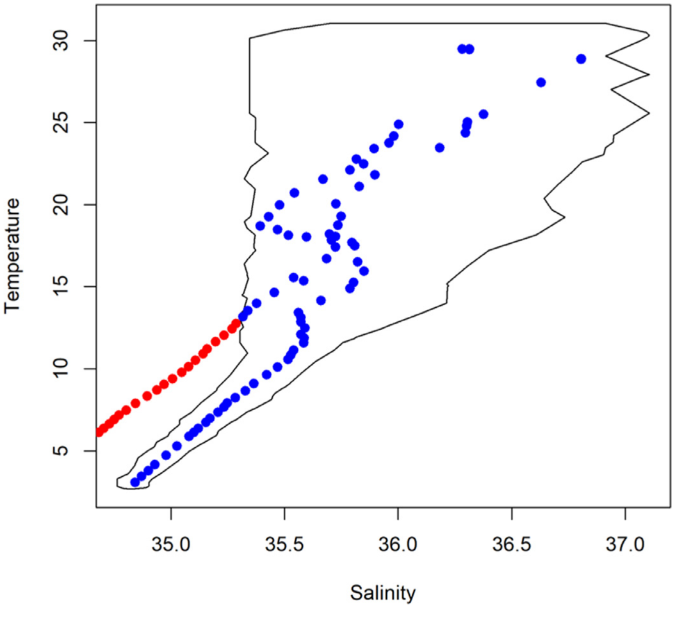

Fig. 3. α shape (convex hull) of annual T/S profiles from WOA13. Good profile falling within α shape is shown in blue color and anomalous profile falling outside α shape is shown in red color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Details of the floats chosen for validation of α convex hull method.

S.No	WMOID	First Cycle	Last Cycle	Total Cycles	Ocean Basin	Type of problem
1.	2900782	22/06/2007	07/01/2012	167	Arabian Sea, Indian Ocean	None
2.	2900877	11/09/2007	04/09/2012	183	Arabian Sea, Indian Ocean	TBTO Fouling
3.	2900554	06/09/2005	27/06/2009	279	Arabian Sea, Indian Ocean	Salinity drift
4.	3900059	12/10/2001	30/09/2002	37	Pacific Ocean	Salinity drift
5.	6901565	05/06/2014	01/05/2019	180	Atlantic Ocean	Salinity drift

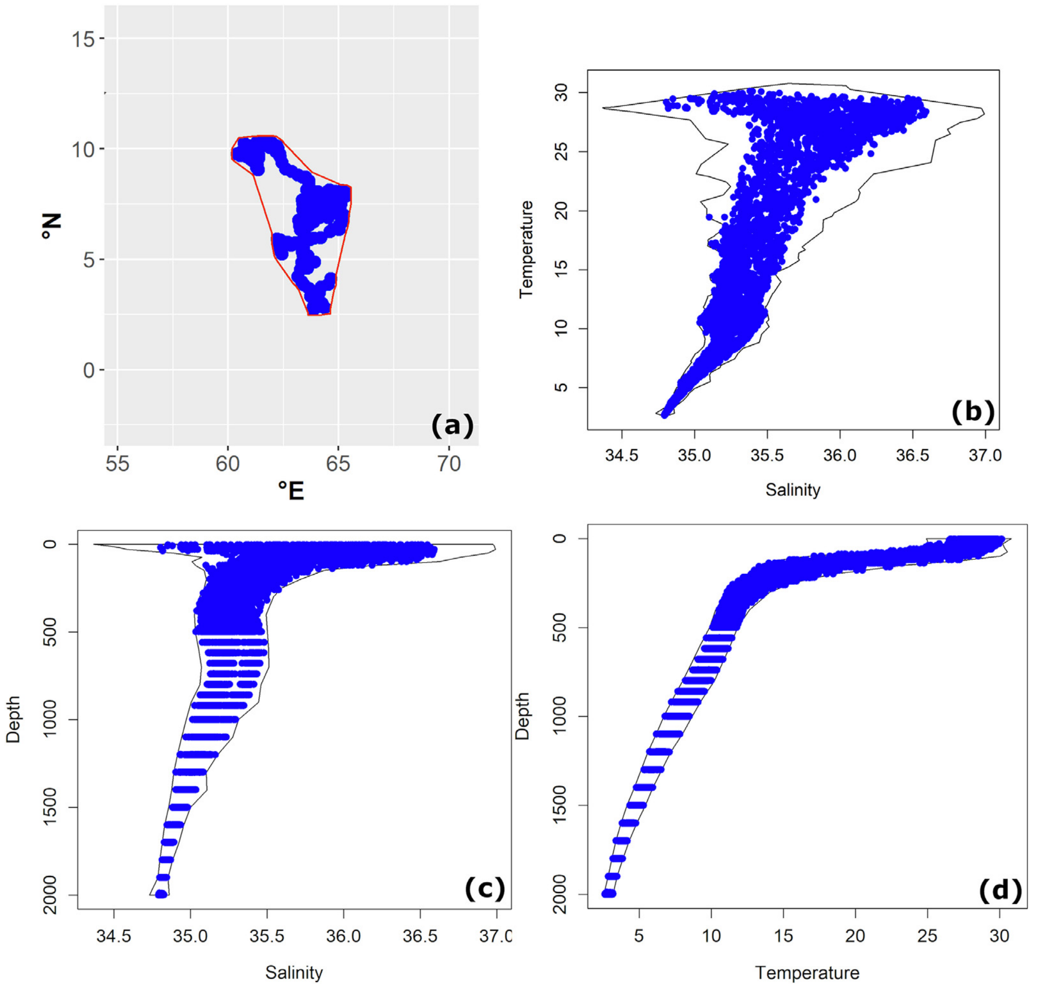

Fig. 4. (a) Trajectory of the float 2900782 during its lifetime. (b) Observed profiles from Argo float overlaid on T/S α convex hull built from annual WOA13 profiles falling within the float trajectory. (c) Observed Salinity profiles from Argo float overlaid on Salinity-Depth α convex hull built from annual WOA13 profiles falling within the float trajectory. (d) Same as that of (c) for Temperature profiles.

(1) The trajectory of observed Argo temperature and salinity profiles are obtained and region of interest (ROI) is defined. ROI is also defined as the optimal convex hull (alpha shape) which contain the trajectory of the Argo float observations.
(2) Use annual World Ocean Atlas (2013) temperature and salinity profile data corresponding to this ROI and build a T/S α convex hull with least area encompassing the mean ± 2*standard deviation fields of temperature and salinity profile fields.
(3) Subsequently the PIP algorithm implemented using Ray Casting Algorithm [6] is used to check if the observed temperature and salinity profile (obtained in step 2) falls within or outside this climatological T/S α convex hull.
(4) Set the quality flags as good(bad) for data falling within(outside) the α convex hull there by identifying erroneous profile data as shown in Fig. 3.

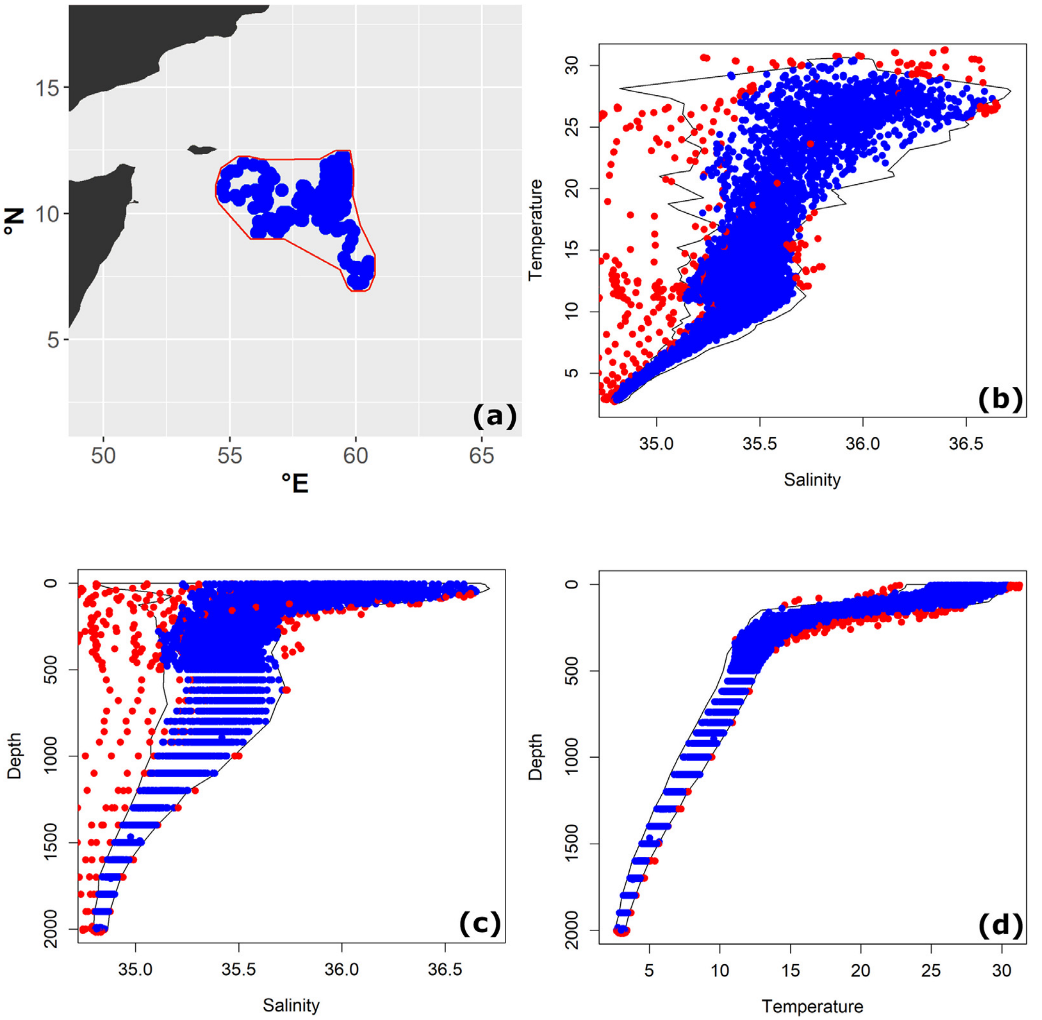

Fig. 5. (a) Trajectory of the float 2900877 during its lifetime. (b) Observed profiles from Argo float overlaid on T/S α convex hull built from annual WOA13 profiles falling within the float trajectory. Blue indicates good and red indicates outlier. (c) Observed Salinity profiles from Argo float overlaid on Salinity-Depth α convex hull built from annual WOA13 profiles falling within the float trajectory. (d) Same as that of (c) for Temperature profiles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The climatology used for the proposed method is the gridded fields of temperature [5] and salinity [12] and their standard deviation fields obtained from World Ocean Atlas 2013 (WOA13) of US National Oceanographic Data center (NODC). These climatological mean and standard deviation fields are then used to build polygons (α convex hulls) of Temperature Vs Salinity, Temperature vs Depth and Salinity Vs Depth. The process of interpolating T/S profiles *in situ* data to standard depths was eliminated, as full length profiles are considered for quality check instead of data corresponding to standard depths, as proposed in earlier work [9]. Jarvis March (1973) algorithm which is also popularly called as gift wrapping algorithm was employed for building these α convex hulls corresponding to months, seasons and annual data sets. The T/S profiles from *in situ* platforms are checked to see if they are falling within the corresponding α convex hull obtained from WOA13 climatology for that

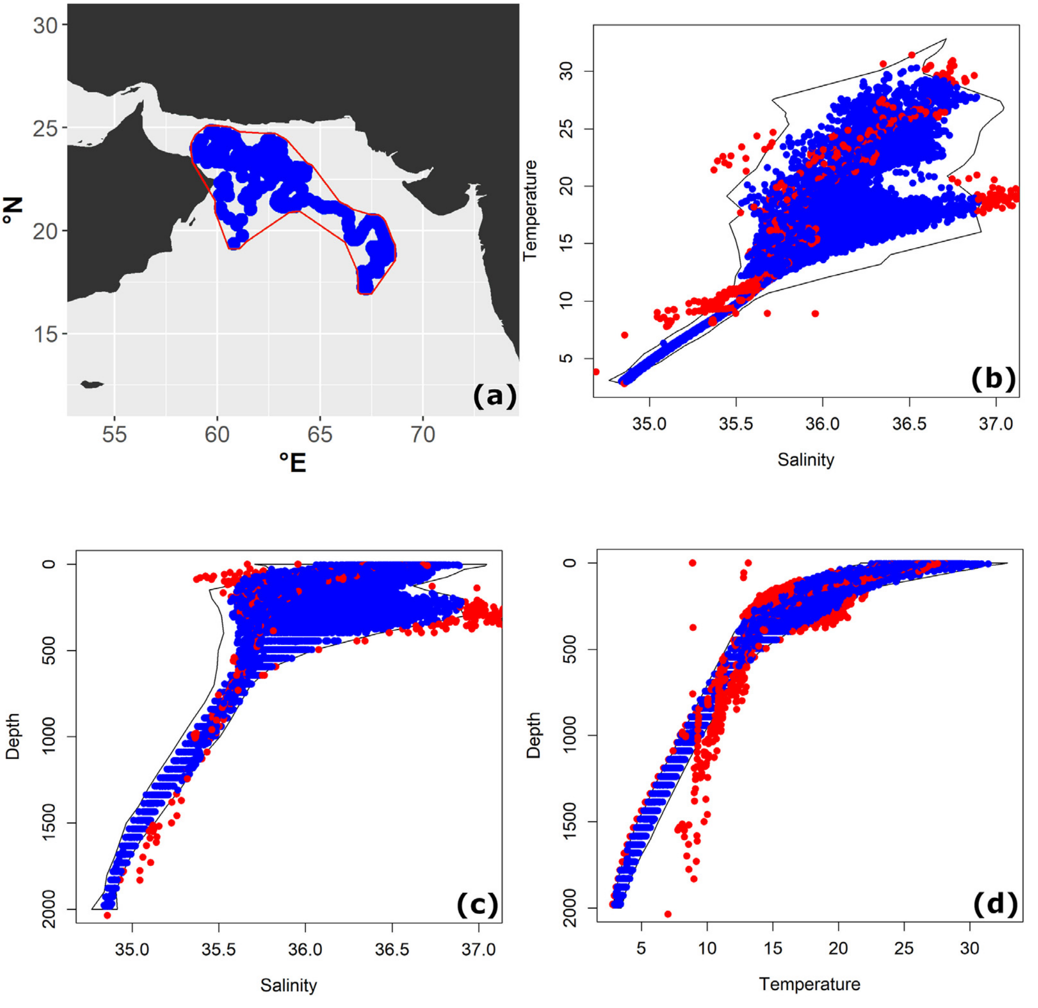

Fig. 6. Same as that of Fig. 5 but for the float with WMO 2900554.

corresponding month, season and annual. PIP algorithm was employed to see if the profile data is falling inside or outside. As with the previous method, this method also has a complexity of O(nh) where n is the number of points and h is the number of points on the α convex hull. For more details on how the algorithm was used for building the α convex hull for individual standard depths, kindly refer to Udaya Bhaskar et al. [9].

Validation of the proposed method

Offsets, freshening due to Tri-Butyl Tin Oxide (TBTO), drift after a set of cycles are some of the well noted problems with salinity sensors. Accordingly the α convex hulls obtained using T/S are used for checking the quality of salinity data in this section. All the climatological temperature and salinity profiles falling within the region encompassing the float trajectory are obtained from WOA13 and α convex hulls are built using the Jarvis March algorithm. Observed Argo float profiles are checked

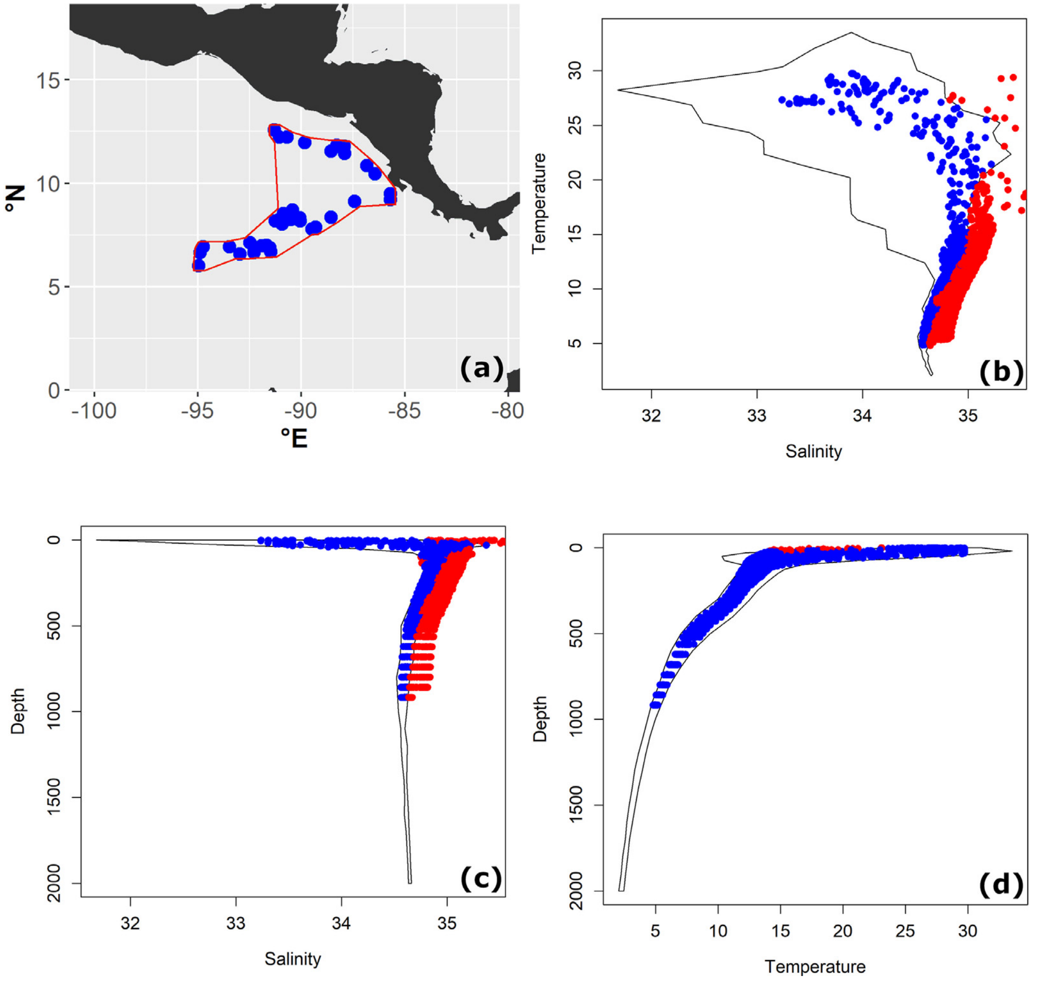

Fig. 7. Same as that of Fig. 5 but for the float with WMO 3900059.

against these n-sided α convex hulls (polygon) using the PIP algorithm. If any of the data points happen to fall outside the α convex hulls, the Argo profiles are suspected to have a problem (drift, bias, spike etc.). Because the climatology is thoroughly quality controlled by experts and has a large number of observations spanning decades, a float is suspected to have a problem if the profiles of temperature and salinity fall outside the n-sided polygon.

To demonstrate the robustness of the proposed method, 5 typical floats are chosen from different ocean basins (Pacific Ocean, Atlantic and Indian Ocean) which represent a float with no problems (good float) and floats with problems like drift, TBTO fouling etc. The details of the floats chosen for the validation are given in the Table 1. To demonstrate the robustness of the proposed method, various anomalous floats were considered.

The first float with WMO 2900782 was deployed in central Arabian Sea and it is observed to be confined to 5° x 5° region and measured 167 T/S profiles in total. This float is typical example of good float which did not have any anomalous observations throughout its life time. Accordingly one can see all the profiles falling inside the n-sided polygon (Fig. 4b), thereby indicating that the float is good and no issues are encountered with the sensors. The salinity profiles are also observed to be fully falling inside the α convex hull built from annual WOA13 salinity profiles (Fig 4c). None of the temperature profiles are observed to be falling outside the α convex hull built from annual WOA13 temperature profiles (Fig 4d). Overall the method proves the float to be good which is matching to those flags set to the data available from Global Data Assembly center.

The second float with WMO 2900877 was deployed in Arabian Sea near to Socotra island and it moved south eastward and measured 183 T/S profiles in total (Fig 5a). This float is observed to be contaminated by TBTO fouling which is clearly evident from the initial few profiles which are represented by fresher salinity values. Sometimes this causes erroneous freshening in the initial profiles until the coating is washed off. Clearly one can see all the initial profiles falling outside the n-sided polygon (Fig. 5b), thereby indicating the case of the TBTO contamination. This is also clearly seen in salinity profiles falling outside the α convex hull build from annual WOA13 salinity profiles (Fig 5c). Few temperature profiles are also observed to be falling outside the α convex hull built from annual WOA13 temperature profiles (Fig 5d).

The third example is the float with WMO 2900554 which was deployed in the northern Arabian Sea north of Oman. This float is also observed to move south eastward and measured 279 T/S profiles during its lifetime (Fig 6a). Its salinity started to drift starting from cycle 200 onwards (Fig 6b) indicated by the red dots falling outside the α convex hull built from WOA13 (Fig 6c). All the salinity profiles corresponding to these cycles are set to quality flag 4 indicating bad quality. The temperature profiles are also observed to be having many outliers (Fig 6d) indicating bad quality.

The fourth example is the float with WMO 3900059 which was deployed in the Pacific Ocean north of Equator. This float is observed to move north eastward and measured 37 T/S profiles during its lifetime (Fig 7a). Its salinity started to drift starting from the fourth cycle onwards and all salinity profiles are observed to be having drift compared to all the nearby buddies. Accordingly all the salinity profiles starting from cycle 5 corresponding to the float are observed to be falling outside the n-sided polygon (Fig. 7c). Very few spurious values corresponding to temperature are observed (Fig 7d).

The last float chosen for testing the method is from the Atlantic ocean with WMO 6901565. This float was deployed northwest of Spain and was confined within 5° longitude, latitude box (Fig 8a) and measured 180 profiles during its lifetime. This float is found to have drift in salinity which can be observed by the salinity profiles falling outside the α convex hull built from annual WOA13 (Fig 8c). Few temperature profiles are also observed to be bad falling outside the α convex hull (Fig 8d).

The sample floats used for testing the proposed method are only an illustration of some possible cases of anomalous behavior of salinity and temperature sensors. This method can also be effectively used to detect spike, offsets in all types of profile data. The method need to be tested thoroughly for complex regions with multiple water masses though initial results showed that the method works hassle free. Further if a floats moves to quite long distance, the homogeneity of the water need to tested and if needed different T/S convex hulls corresponding to different water masses/basins need to derived and tested for those sets of profiles falling in that region. This new/improved method is definitely a significant improvement over the originally proposed method by Udaya Bhaskar et al. [9] as it treats the whole observed profile and eliminates the possible error that might creep in due to interpolation. The applicability of the method for different ocean basins demonstrate the usefulness of the proposed method for identifying bad profiles over the global ocean to the extent possible thereby minimizing manual invention by experts. The big advantage of using the α convex hull method is its applicability to handle bulk amounts of profile data from any oceanic basis. As it is nearly impossible for experts to visually check the correctness of individual profiles, this method can be used to detect good against bad profiles. The profiles rejected by this method can be passed through visual inspection by experts making their task simpler and better. However this method needs climatology to be updated continuously for obtaining the best possible α convex hulls for performing outlier analysis. Further, this method can be augmented with other methods in use by the oceanographic community to make the data research quality and to be used for various applications.

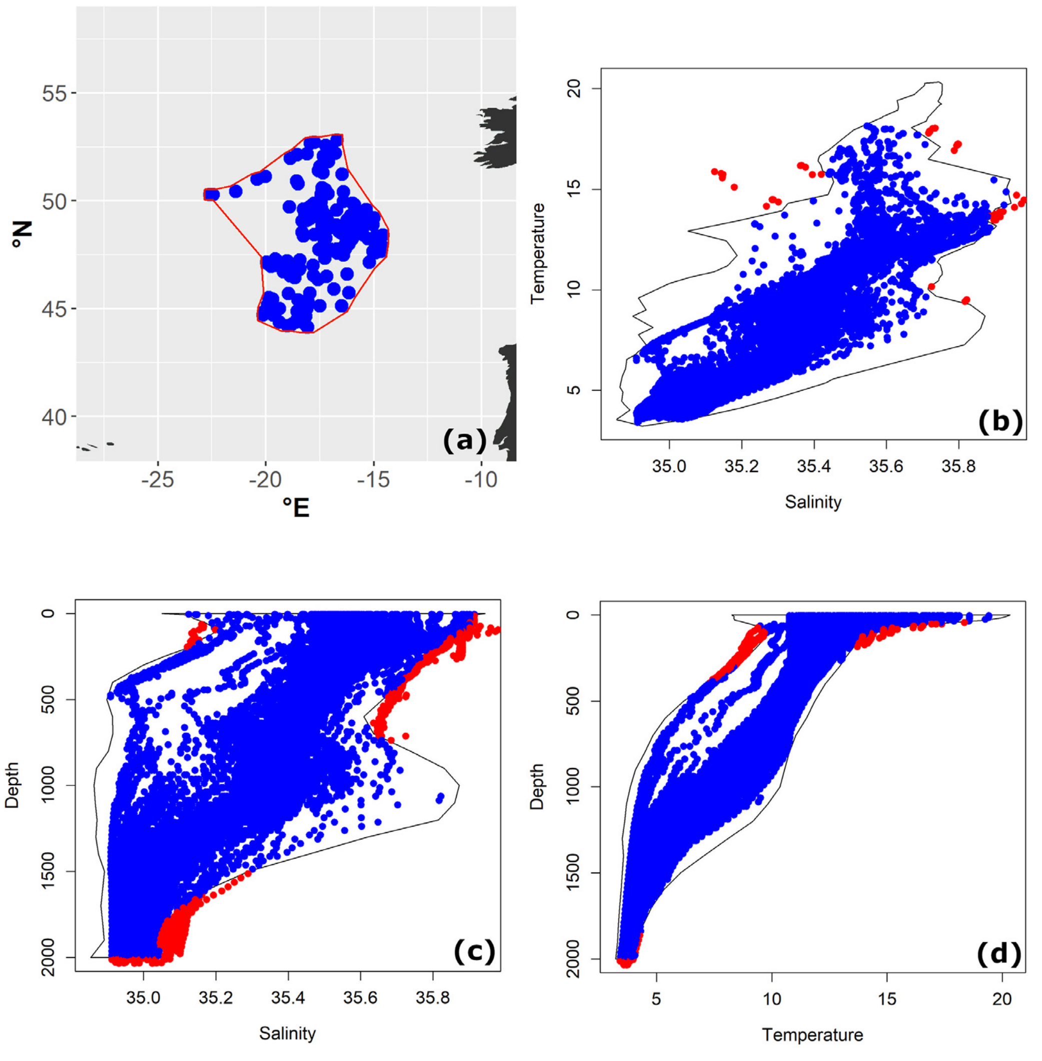

Fig. 8. Same as that of Fig. 5 but for the float with WMO 6901565.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

Authors thank Director, INCOIS for his support to carry out this work. Argo data (http://doi.org/10.17882/42182) 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. This is INCOIS contribution number 413.

References

[1] A. Chatterjee, D. Shankar, S.S.C. Shenoi, G.V. Reddy, G.S. Michael, M. Ravichandran, V.V. Gopalkrishna, E.P. Rama Rao, T.V.S. Udaya Bhaskar, V.N. Sanjeevan, A new atlas of temperature and salinity for the North Indian Ocean, J. Earth Syst. Sci. 121 (3) (2012) 559–593 June 2012.

[2] H. Edelsbrunner, D.G. Kirkpatrick, R. Seidel, On the shape of a set of points in the plane, IEEE Trans. Inform. Theory 29 (4) (1983) 551 {559. ISSN 0018-9448.

[3] W. Emery, J. Dewar, Mean temperature-salinity, salinity-depth and temperature-depth curves for the North Atlantic and the North Pacific, Prog. Oceanogr. 11 (1982) 219–305 Pergamon.

[4] R.A. Jarvis, On the identification of the convex hull of a finite set of points in the plane, Inf. Process. Lett. 2 (1973) 18–21, doi:10.1016/0020-0190(73)90020-3.

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

[6] M. Shimrat, in: Algorithm 112, Position of Point Relative to Polygon, CACM, 1962, p. 434.

[7] P. Thadathil, P.M. Muraleedharan, V.V. Gopalakrishna, G.V. Reddy, L. Ratnakaran, C. Revichandran, and V.S.N. Murthy, Validation of ARGO data in the Indian Ocean, Gayana, 68(2) (2004) 456–458.

[8] Udaya Bhaskar, T.V.S., M. Ravichandran, R. Devender (2007) An operational objective analysis system at INCOIS for generation of Argo value added products, Tech. Rept, INCOIS-MOG- ARGO-TR-04-2007.

[9] T.V.S. Udaya Bhaskar, R.Venkata Shesu, T.P. Boyer, E. Pattabhi Rama Rao, Quality control of oceanographic *in situ* data from Argo floats using climatological convex hulls, MethodsX 4 (2017) 469–479.

[10] A.P.S. Wong, G.C. Johnson, W.B. Owens, Delayed-mode calibration of autonomous CTD profiling float salinity data by theta-S climatology, J. Atmos. Ocean. Technol. 20 (2) (2003) 308–318.

[11] Worthington, L. (1981) The Water Masses of the World Ocean: Some Results of Fine-Scale Census. Evolution of Physical Oceanography: Scientific Surveys in Honor of Henry Stommel, B. A. Warren and C. Wunsch, Eds., The MIT Press, 42–69.

[12] Zweng, M.M., J.R. Reagan, J.I. Antonov, R.A. Locarnini, A.V. Mishonov, T.P. Boyer, H.E. Garcia, O.K. Baranova, D.R. Johnson, D. Seidov, M.M. Biddle (2013) World Ocean Atlas 2013, Volume 2: Salinity. S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 74, pp. 39