Quality control method of Argo float position data

Tomoaki Nakamura[1], Naoko Ogita[1] and Taiyo Kobayashi[1]

Abstract We explain a method of quality-control (QC) of Argo float positions fixed by Argos system based on checking float speed between the positions considering Argos position errors. The method gives us fairly reasonable QC results which are comparable with those of visual inspection by experts, and several % of position data are identified "bad" in average. This method has been suggested as (one of candidates for) the standard method to International Argo community. An execution program of it was also prepared for Argo community from PARC-JAMSTEC web-site so as that this method is tried widely.

Keywords: Argo program, float trajectory, velocity, quality control

1. Introduction

This manuscript explains a method of quality-control (QC) of Argo float positions fixed by Argos system based on check of float speed between the positions. This document also describes some results obtained by this QC scheme and the QC method has been suggested as a standard method to International Argo community.

We suppose that this position QC procedure is the first check for Argo trajectory data to discard (or flag) very "strange" positions automatically. We expect (and it has been discussed in Argo community) that some of more sophisticated interpolating schemes (e.g., Davis et al., 1992[3]), Park et al., 2005[5]), "delayed-mode QC" for Argo trajectory data) will succeed it to estimate actual float movements and locations where floats arrive at sea surface and dive into subsurface. Since "strange" positions distort these estimations largely. On the other hand, the more data we can use for the interpolations the better results we can have. Thus, it is desirable that this position QC scheme discards only very "unusual" position data for most of experts.

Another purpose of the suggestion of this QC method is to standardize data quality of Argo trajectory data by a single QC scheme. Trajectory files are generally used to estimate velocities on the sea surface and/or at the parking depths of floats directly. Large amount of data are needed regardless of float principal investigators (PIs) or Data Assembly Centers (DACs) when basin (global) scale velocity fields are calculated (e.g., YoMaHa dataset, Konstantin et al., 2007[4]). Quality of such velocity atlases may depend on the lowest level of original data. Thus, standardized quality of Argo trajectory data brings lots of benefits to all data users.

To standardize the data quality we believe a single standard program is needed definitely considering the present confusion of Real-time QC of float profiles. We know that its results by DACs are very different from each other even though they do follow the standard QC criteria defined by Argo Program completely. It means that only definitions of criteria (like the tables of Real-time QC flags for float profile, Argo Data Management Team, 2006[1]) might not be enough to dissolve differences among DACs in the case of complicated data like float profiles and float trajectories.

First, we described backgrounds of this QC method development in section 2 in this manuscript. Then the actual procedure of this scheme is mentioned (section 3). In section 4 some results of position QC are shown and an executive program of the QC method is introduced in section 5. Finally, we mentioned some issues to be discussed or agreed in Argo community in section 6.

2. Background

Argo program aims not only to monitor the present status of the upper ocean (up to 2000-dbar) from float temperature and salinity profiles but also to estimate velocity fields on the sea surface and at float parking depths from float drifting (trajectory) data. However, until now there are only a few studies about the velocity field estimations from the Argo data (e.g., Davis et al., 1992[3]), Park et al.,

1 Institute of Observational Research for Global Change, Japan Agency for Marine-Earth Science and Technology

2005[5]), Konstantin et al., 2007[4]). One of the reasons for it may be that Argo trajectory data (including float position data) are not as familiar as the Argo profile data which have temperature/salinity profiles. But, the most reason we think (and the most of persons involved in Argo will agree) is that Argo trajectory data are very far from complicity (or no user-friendly), because its detail manners of data description are not (have not been) determined well to use them. This undesirable situation may be caused by the fact that the most of human resources of Argo (especially in its data management) are being consumed in development/maintenance of data system for Argo profiles.

In order to improve these situations around Argo trajectory data, which was also requested by Argo Steering Team, the first and second Argo trajectory workshop was held in March 2006 (in Venice, Italy) and in October 2006 (in Incheon, S. Korea). Both workshops were led by Dr. B. A. King. At the second workshop one of the authors (TK) introduced a position QC method used at JAMSTEC. The method was highly evaluated there and it was recognized as a candidate for the standard QC method of Argo trajectory (position) data for the following reasons;

– In Argo community there were almost no determinations (definitions) about the details of Argo trajectory data. Thus, they wanted to have something concrete about them.
– This QC scheme was actually used at JAMSTEC, which means its source code was available in Argo community. This is one of large benefits for the community considering the then complicated problem on Real-time QC of Argo profiles, where QC results by

DACs were different from each other even though its algorisms and criteria have been defined clearly (and were followed by all DACs).

At the same time, it was requested that this QC method should be improved to consider position errors fixed by Argos system. The position QC method described in this manuscript is the modified scheme considering the comments at the workshop.

3. Position QC procedure

This position QC scheme identifies "bad" float positions considering the float speeds which are required by the float positions fixed by Argos system.

3.1. Identification of "questionable" segments

Assuming a float movement consists of numerous segments of two adjacent positions in time. If a segment which satisfies the following 2 criteria (a "questionable" segment, see Fig. 1) is found, at least one position which composes the segment will be flagged 3 ("bad").

– Float speed along the segment exceeds 3 m/sec.
  • The criterion of 3 m/sec for "questionable" float speed is derived from a criterion of surface movement in "Argo quality control manual" (Argo Data Management Team, 2006[1]).
  • The minimum time interval is set 1 second to avoid the calculation error of "dividing by 0".

– The length of the segment is longer than the critical error length which is determined by the nominal errors of Argos system as follows:

$$\text{the critical error length} = 1.0 \times \sqrt{Er_A^2 + Er_B^2}.$$
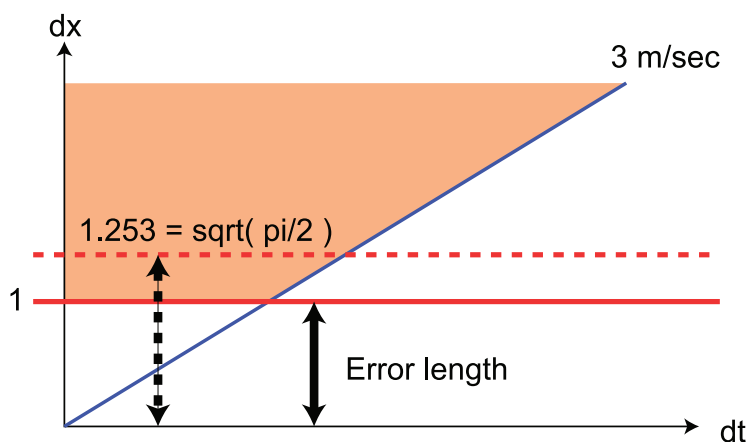

Figure 1: Definition of "questionable" segment which includes at least one float position to be flagged 3 (hatched region). dx and dt are horizontal distance and time interval between 2 float positions fixed by Argos system. Error length is determined by $1.0 \times \sqrt{Er_A^2 + Er_B^2}$, Er_A and Er_B are the radii of position error of Argos system (150m, 350m, and 1000m for Argos class 3, 2, and 1, respectively) at the float positions A and B, respectively.

Here, Er_A and Er_B are the radii of position error of Argos system (150m, 350m, and 1000m for Argos class 3, 2, and 1, respectively) at the float positions A and B, respectively. The factor of 1.0 above is one of optional choices (see section 6.2).

Fig. 2 shows scatter plots of horizontal distance and time interval between any two positions in the same cycle of a float (n = 4379 pairs for 290 float-cycles). The criterion of 3 m/sec for float speed is not severe except for the position pairs with shorter time interval (about less than 10 minutes). Only in this time range the Argos positioning errors can not be ignored for float speed estimation. It means that the above criteria are reasonable.

3.2. Identification of "bad (flagged)" position

"Bad (flagged)" position of the questionable segment is identified by the following steps:
– If Argos classes at the positions are different, the position with the less grade class is flagged.

Accuracy of Argos class:

More accurate <= 3, 2, 1, 0, A, B => Less accurate
Here, the position with Argos class 0 means its error radius is grater than 1000m, and Argos class A (B) represents that the float position is fixed by 3 (2) messages received per satellite pass and its accuracy is not officially determined (positions with Argos class 3, 2, 1, and 0 are determined by 4 or more messages and more accurate than
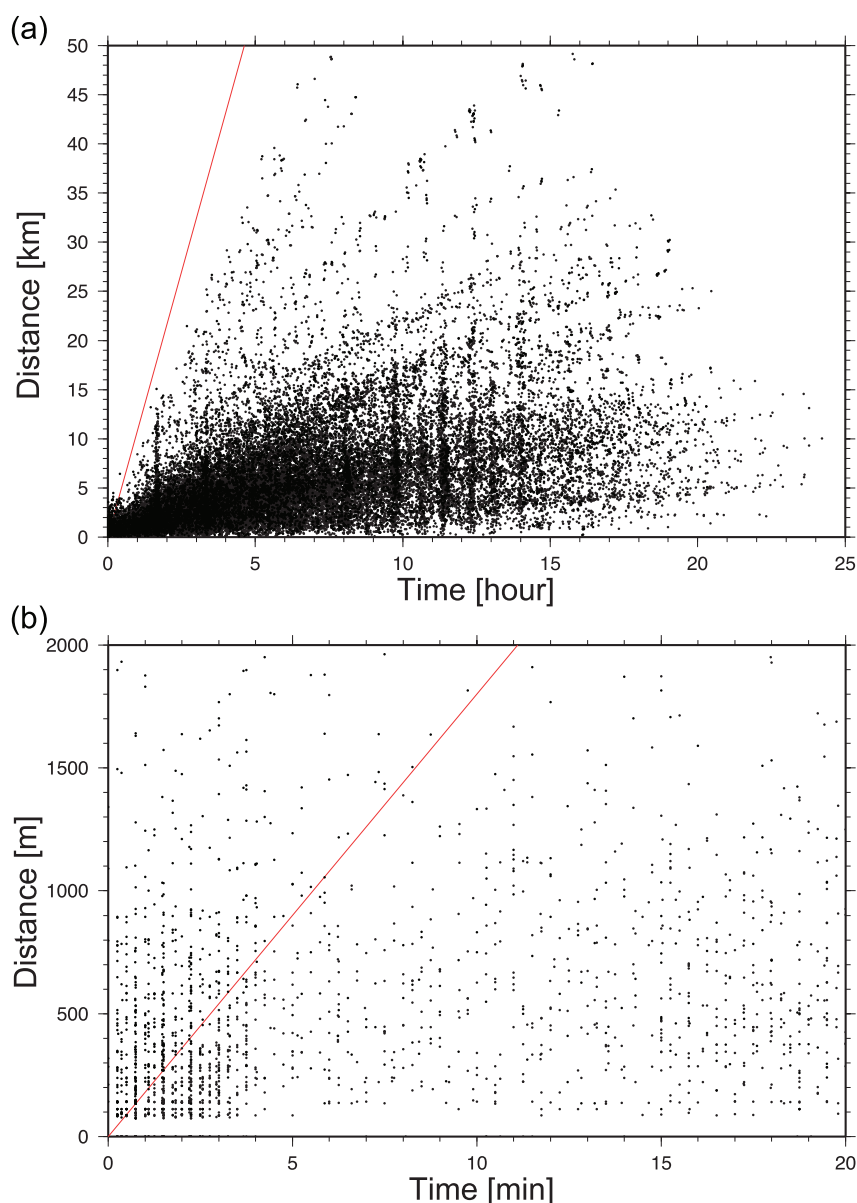

Figure 2: Scatter plots of horizontal distance and time interval between any two float positions in the same cycle (n = 4379 for 290 float-cycles). The red line represents the relation of float movement by 3 m/sec.

those with class A and B: CLS, 2007[2]).
– If Argos classes at both positions are the same, the position which requires more quick movement along the continuous segments is flagged (see also Fig. 3).

- In the case that 4 continuous positions can be used for the check, average speeds along the segments via A and via B are compared (yellow and cyan trajectories in Fig. 3a).
- In the case that 3 continuous positions are available for the check (i.e., the questionable segment includes the first/last position of the cycle), we compare average speeds on the segments terminated at A and at B (yellow and cyan trajectories in Fig. 3b).
- If we use only 2 positions for the check (i.e., the float cycle for check is composed of only 2 positions), both positions are flagged 3 because we can not determine which position is worse than the other (an exceptional case).

3.3. Other "bad" data identified in this scheme

Some of Argo trajectory data we have checked included very strange position data and we concluded that they should be flagged 4. Because most of them should be removed initially by other tests before the QC method discussed here.
– Duplication of position data (both location and time).
   => They are flagged 4 except for a single data. If different positions are fixed in the same time, all positions are checked by the QC scheme.
– Positions which are apart from the first position by more than 24 hours in time.
   => It is considered that surface-drifting data of another cycle may be involved.

– Float speed in subsurface (from the last "reasonable" position of the previous cycle to the first position of the present cycle) exceeds 3 m/sec.
   => Questionable float position due to too fast movement in subsurface.
– Unrealistic dates of float positions fixed for Argo trajectory data.
   => Adjusted to the "right" dates.
– Values for Argos class are written by "octal notation".
   => Adjusted to the decimal notation.

4. Results

Fig. 4 shows some of position QC results obtained by the method. Fig. 4a is results for a float deployed in the subarctic region of the North Pacific (77th cycle of WMO ID 2900055). The first position of this cycle was fixed on 12:03, 22 July 2003 (Argos class 1), and then the float drifted north-eastward generally. The red line represents a float trajectory obtained by connection of float positions which passed through the QC scheme. This shows that the float moved sometimes zigzag, and some of them seem unrealistic movements. But all the positions except for one (on 00:08, 23 July 2003, represented by a red star) are fallen within error radii of float positions fixed by Argos system.

Fig. 4b shows another example (54th cycle of WMO ID 2900056). This float drifted on the sea surface within very small area and most of the circles of positioning errors overlapped each other. The float may have moved north-eastward slowly in the cycle when positioning errors are considered, and 2 positions are failed by the QC check.

Fig. 4c is a result that many position data are failed. In this float-cycle total 19 positions were fixed by Argos sys-
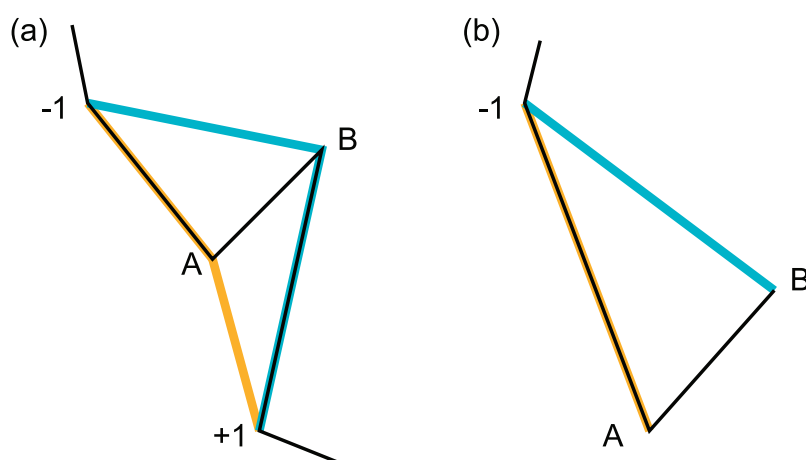

Figure 3: Schematic figures of identification of the position to be flagged in the cases of (a) 2 next positions (signed by -1 and +1) of the "questionable" segment (AB) and (b) only one next position (-1, the previous or next) of the segment can be used for the check.
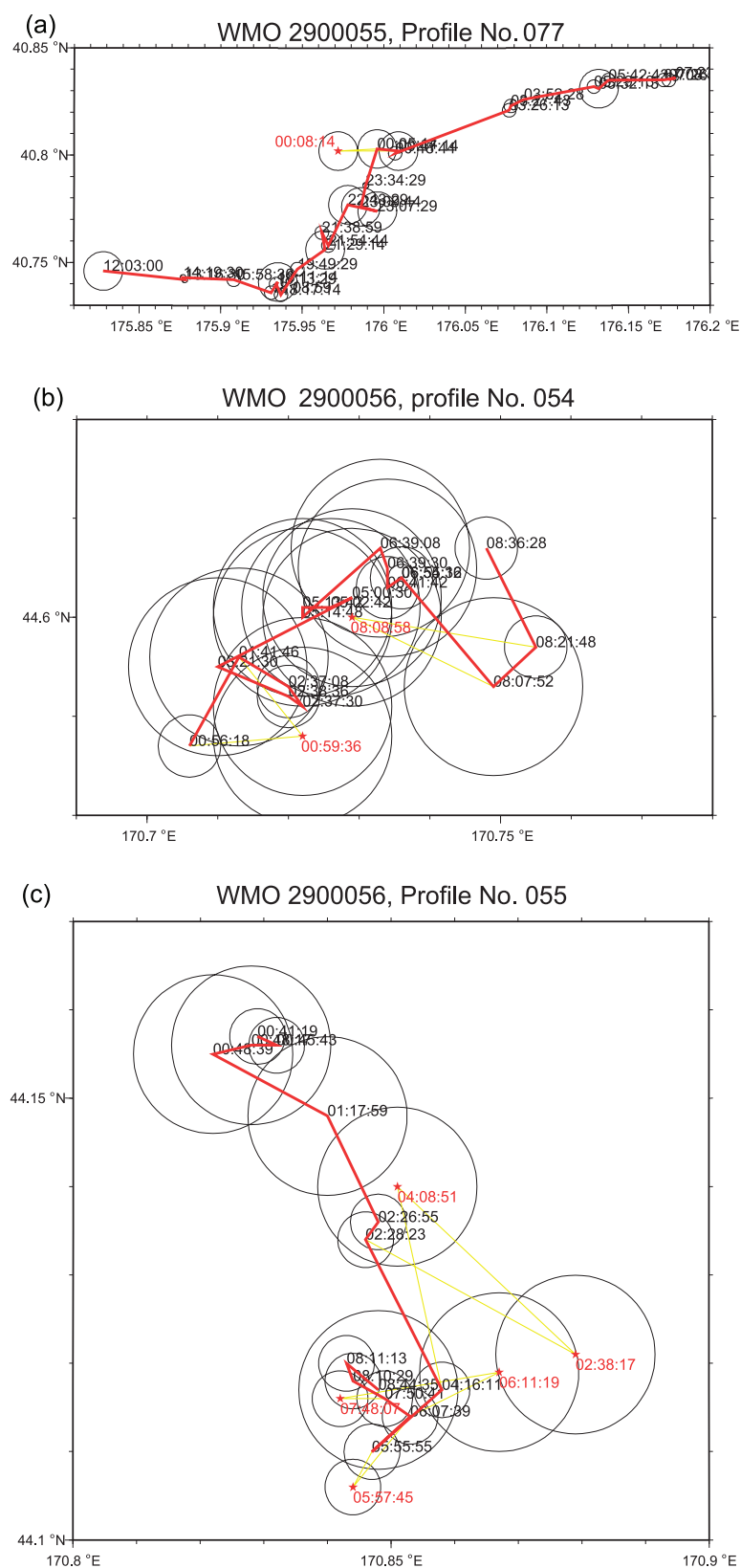
Figure 4: Examples of the position QC results for (a) the 77th cycle of WMO ID 2900055, (b) the 54th cycle and (c) the 55th cycle of WMO ID 2900056. The flagged positions are shown by red stars, and positions though the QC are connected by red line in time. The time when the position is fixed by Argos satellites is also shown. Circles mean error radii of Argos system for the positions (1000m, 350m, and 150m for Argos class 1, 2, and 3, respectively).

tem and 5 data of them were flagged 3 by this QC scheme. The 8th and 9th position in the cycle (fixed on 2:38 and 4:08, 16 December 2002) was very far away from the other succeeding (7th and 10th) position considering the time fixed, which shows they are erroneous positions clearly.

Table 1 is the summary of position QC results for 26 Argo floats (for their whole lives) and they are randomly selected from GDAC. This QC method identifies several % of position data to be flagged 3 (or 4) in average. This ratio varies largely (about 1 % to more than 10%) by float and PI (or DAC). The differences among DACs are mainly caused by features of their trajectory files (e.g., positions with less accurate Argos classes (0/A/B) are included or not).

5. Execution program of the position QC method for Windows

We have prepared an execution program of this QC method for Windows so as that this method should be tried widely in Argo community. This program and everything related to it are available from the page of "Tools and Links" at PARC-JAMSTEC web-site.

PARC-JAMSTEC: http://www.jamstec.go.jp/ARGORC/

This program reads netCDF files of Argo trajectory in direct and then outputs QC flags of float positions by text file. Also it creates a KML file of the results for "Google Earth", which helps us to check results easily by visual inspection (see Fig. 5 as an example). Some parameters of the QC method (the maximum speed, positioning error radius of Argos system) are changeable by its configuration file (postion_qc.cfg).

If you would like to modify this program largely, its source codes are also available from the same web site. Note that the source codes are written by "Ruby".

6. Issues to be discussed in Argo community

In developing the QC system, we found several issues which should be discussed and agreed among the Argo community. Some of them may need changes of parameter definitions.

6.1. Argos class and its error radius

Argos class represents the accuracy of a float position fixed by Argos satellites and it has 7 categories of 3, 2, 1, 0, A, B and Z. Except for Z, which means position fixing error, a location data has a corresponding Argos class. However, the accuracy (error radius) of Argos class for 0, A, and B is not determined officially. In this QC system positions with all Argos classes can be examined and those error radii for Argos class of 0, A and B are set as 1500m uniformly (default value: it is changeable in the system). International standardization of the position QC method needs some agreement about

Table 1: Ratio of flagged positions by Argo float. Ratio is evaluated by (the number of position data flagged 3 or 4 by the QC method) / (the number of position data during the float life).

WMO ID	DAC	Ratio
4900369	AOML	1.10%
4900598	AOML	4.20%
6900091	AOML	1.80%
1900097	BODC	4.60%
1900288	BODC	1.20%
1900453	BODC	1.60%
2900194	BODC	6.00%
6900082	BODC	9.10%
6900083	BODC	9.80%
6900193	BODC	7.30%
53553	CSIRO	0.70%
2900425	ifremer	4.90%
3900137	ifremer	1.40%
3900195	ifremer	3.10%
6900139	ifremer	10.50%
6900279	ifremer	1.70%
7900097	ifremer	2.20%
2900342	INCOIS	13.00%
2900343	INCOIS	12.10%
2900354	INCOIS	10.60%
2900463	INCOIS	8.40%
2900287	JMA	5.50%
2900312	KMA	2.80%
4900101	MEDS	5.60%
4900414	MEDS	3.90%
4900517	MEDS	1.20%

Argos class to be used for position QC and their error radii. This is closely related to creation of Argo trajectory data, e.g., what float positions should be recorded there.

6.2 Error length criterion

The length criterion used in the method is much simplified to be factor 1, by which our concept is understood easily. The preliminary statistic examinations gave us another candidate for the factor,

$$1.253 \left(= \sqrt{\pi / 2}\right)$$ (see Figs. 1 and 6).

6.3. Consistent description of Argo trajectory data

This QC scheme needs to deal with "strange" data in Argo trajectory files. It is better that they are removed before the position QC, which gives us much simpler codes of the position QC scheme and also lots of benefits to all persons who use Argo trajectory data directly. We suggest that the trajectory data which satisfy the following criteria be required for any position QC method.
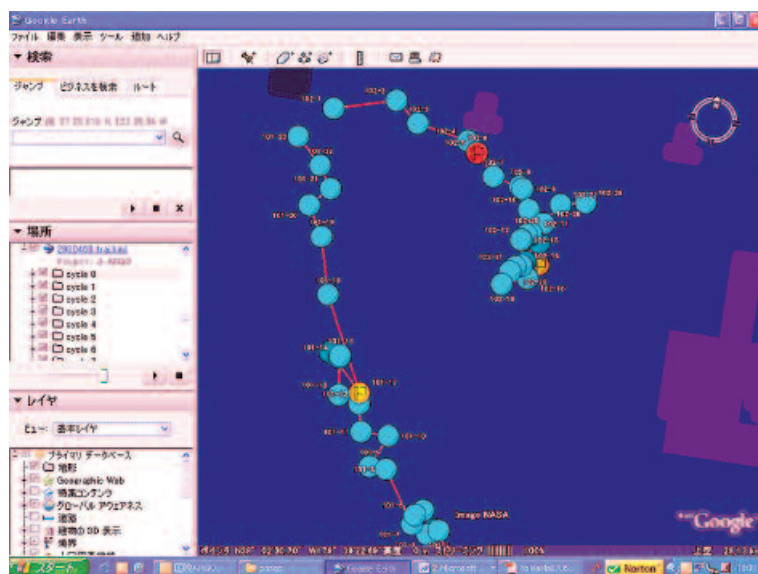
Figure 5: An example of output of position QC results for Google Earth. Yellow and red marks represent the positions with flag 3 and 4, respectively.
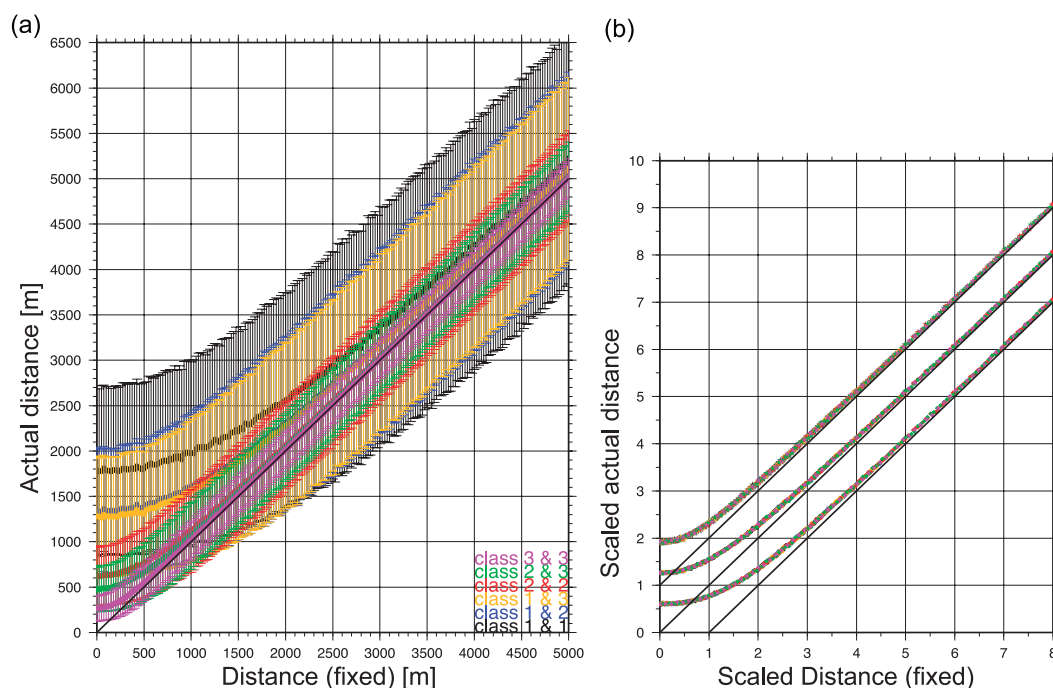

Figure 6: Estimation of "actual" distance between float positions fixed by Argos system with Monte Carlo method. (a) Average and standard deviation of the actual distance between two positions (as a function of the distance fixed by Argos) considering Argos error distance. Color means each combination of float positions with different Argos class. (b) Average +/- standard deviation of the distances normalized by the factor of $\sqrt{Er_A^2 + Er_B^2}$.

– All data are filled by the format determined officially.
– All data (e.g., date, longitude, latitude …) fall in reasonable values.
– All data are ascending ordered in time.
– No complete duplications (all data of date, position, Argos class are the same).
– All positions have corresponding data of Argos class.
– No large (temporal) gaps in the same cycle, because such the data can be considered as miss involving of another cycle data.

References

1) Argo Data Management Team, "Argo quality control manual version 2.2," 33pp. (2006).
2) CLS, "Argos user's manual," (2007).
3) Devis, R.E., D. C. Webb, L. A. Regier, and J. Dufour, "The autonomous Lagrangian Circulation Explorer (ALACE)," *J. Atmos. Oceanic Technol.*, 9, 264-285 (1992).
4) Konstantin V. Lebedev, Hiroshi Yoshinari, Nikolai A. Maximenko, and Peter W. Hacker, "YoMaHa'07: Velocity data assessed from trajectories of Argo floats at parking level and at the sea surface," *IPRC Technical Note*, 4(2) 16pp. (2007).
5) Park, J. J., K. Kim, B. A. King, and S. C. Riser, "An advanced method to estimate deep currents from profiling floats," *J. Atmos. Oceanic Technol.*, 22, 1294-1304 (2005).