Technical Report Report No. : INCOIS-DMG-TR-2012-07

A three way semi-automated approach for quality control of Argo temperature and salinity profiles

TVS Udaya Bhaskar, E Pattabhi Rama Rao, R Venkat Shesu and R Devender

Indian National Center for Ocean Information Services
(Ministry of Earth Sciences, Govt. of India)
Hyderabad

August, 2012

DOCUMENT CONTROL SHEET

---

01. Report No: **INCOIS-DMG-TR-2012-07** Date: **08 August 2012**

---

02. Title & Sub Title:

A three way semi-automated approach for quality control of Argo temperature and salinity profiles.

---

03. Part No.: — 04. Vol. No.: —

---

05. Author(s):

TVS Udaya Bhaskar, E Pattabhi Rama Rao, R Venkat Shesu, R Devender

---

06. Originating agency (Group/Project/Entity): **DMG**

---

07. No. of Pages: **15** 08. No. of figures: **3**

---

09. No. of references: **8** 10. No. of enclosures/appendices: **2**

---

11. Abstract (Maximum 100 words):

A three way semi-automated quality control system for quality control of temperature and salinity profiles obtained from Argo floats is presented. In the beginning, all the temperature and salinity profiles are passed through 18 automated quality checks as suggested by the International Argo Data Management Team. Further, all the profiles are utilized in generating objectively analysed product. Bad profiles appearing as bulls eye are automatically rejected based on preset statistics. These bad profiles are then visually checked (which requires manual intervention) using a visual quality control tool developed in house for their correctness. At the end of the quality control procedure, monthly objective analysis is produced as a by product, which is made available on INCOIS Live Access Server in NetCDF format.

---

12. Keywords: **Argo, Objective Analysis, Visual QC, RTQC**

---

13. Security classification: **Unrestricted**

Table of Contents

Abstract

A three way semi-automated quality control system for quality control of temperature and salinity profiles obtained from Argo floats is presented. In the beginning, all the temperature and salinity profiles are passed through 18 automated quality checks as suggested by the International Argo Data Management Team. Further, all the profiles are utilized in generating objectively analysed product. Bad profiles appearing as bulls eye are automatically rejected based on preset statistics. These bad profiles are then visually checked (which requires manual intervention) using a visual quality control tool developed in house for their correctness. At the end of the quality control procedure, monthly objective analysis is produced as a byproduct, which is made available on INCOIS Live Access Server in NetCDF format.

1. Introduction

Argo is an internationally coordinated program directed at deploying the global ocean with 3000 profiling floats that measure temperature and salinity. The profiling float sinks after launch to a prescribed pressure level, typically 2000 dbar and after a preset time (typically 10 days) the float returns to the surface, collecting T/S profiles (Argo Science Team, 2001; Ravichandran et al., 2004). On the surface the float transmits the data to ARGOS satellite which are received at the ground station and are analyzed.

Since the year 2000, the number of Argo floats in the world oceans has been increasing year by year and the projects goal was reached in mid 2008, i.e., more than 3000 floats are now in operation in the world oceans. The number of profiles obtained annually by Argo in the world oceans was more than 30,000 in 2003 and this number tripled to about 90,000 in 2006. Thus, the annual total of Argo profiles obtained every year is now equivalent to three quarters of the total number of historical Conductivity-Temperature-Depth (CTD) profiles deeper than 2000 m archived in the world ocean data base 2001.

Quality control (QC) is an important part of ocean data assimilation system. If erroneous values are assimilated they can cause immediate spurious overturning and also error propagation due to the sparse distribution of oceanographic data (Bruce et al., 2007). In preparing his ocean atlas Levitus (1982) noted that ñBy far the biggest problem faced in this project concerned quality control of the dataò. Various data centers have developed different QC procedures and also documented the same. The semi-automated system described here processes temperature and salinity (T/S) profiles obtained from Argo profiling floats, and with minor variation it is suitable for both real-time and archive

CTD and other sources of data for global and/or regional domain. The output of this QC process is specifically targeted for generation of objectively analyzed gridded product which can be used for studying the variability of the ocean over a period for which the gridded product is available. Further, this gridded product can also be used for assimilation into ocean models.

Taking in to view the vast number of T/S profiles produced by Argo floats, manual quality control of all these profiles is not affordable and time consuming. In this work, we present a semi-automated three way quality control system developed at INCOIS for near-real time quality control of the Argo T/S profiles. These methods use a combination of automated QC procedures (real time checks and Objective Analysis) and visual QC system (which require manual intervention) which increase the quality and reliability of the data there by increasing the value of these important data sets for climate and ocean studies and reanalysis efforts.

The rest of the report is organized as follows. Section 2 briefly reviews the data and methods and describes the QC processing and checks which are relevant to these observations. Section 3 provides a discussion and summary.

2. Data and Methods

Indian National Centre for Ocean Information Services (INCOIS) being the National Data Centre (DAC) for India and Argo Regional Centre (ARC) for the Indian Ocean, archives Argo T/S data from Indian Ocean (20° ï 140° E and 70° S ï 30° N). Profiles obtained from various floats deployed by participating countries are archived at INCOIS on a day to day basis. Each DAC at various countries is responsible for the real time quality control (RTQC) of floats deployed by their respective countries as prescribed by the Argo Data Management Team (Wong et al.,2012). In the light of this, INCOIS performs the automated real time QC of 254 floats deployed by India. Apart from this all the T/S profiles archived at INCOIS undergo additional QC processes. These additional QC procedures along with the automatic RTQC procedures forms the three way QC system which is implemented at INCOIS.

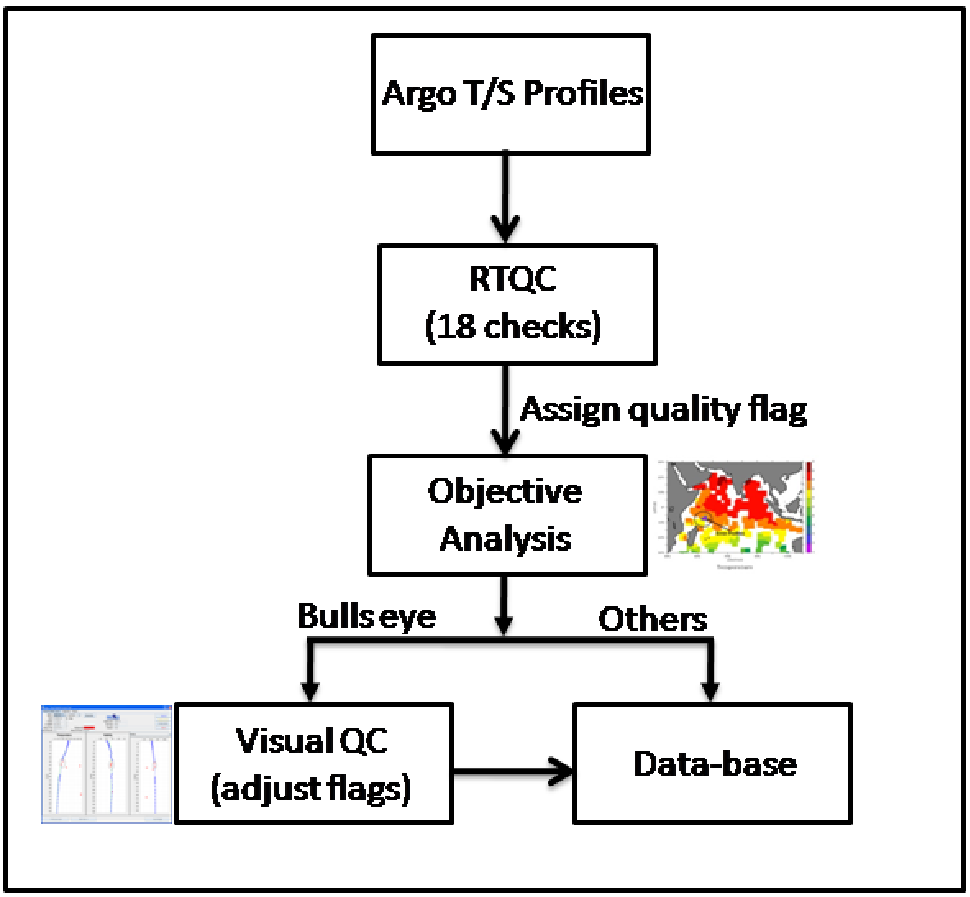

Figure1: Flow diagram showing the various quality control procedures implemented at INCOIS for assigning the quality to CTD data.

1. Platform identification	**2. Impossible date test**
3. Impossible location test	**4. Position on land test**
5. Impossible speed test	**6. Global range test**
7. Regional range test	**8. Pressure increasing test**
9. Spike test	**10. Top and bottom spike test**
11. Gradient test	**12. Digit rollover test**
13. Stuck value test	**14. Density inversion**
15. Grey list	**16. Gross salinity or temperature sensor drift**
17. Frozen profile test	**18. Deepest pressure test**

Table1: Real Time Quality Control (RTQC) tests as prescribed by the Argo Data Management Teal (Wong et al., 2012).

Figure 1 shows the flow of the three way quality control system set up at INCOIS. The CTD profiles obtained from Argo floats is subjected to three levels of QC as described below:

- **Automatic QC:** The first level is the real-time system that performs a set of agreed checks (as prescribed by Argo Data Management Team) on all float measurements. Real-time data with assigned quality flags are available to users within the 24 hrs timeframe.
- **Objective Analysis:** The second level of QC involves usage of objective analysis for identification of outlier which appear as bulls eyes.

- **Visual Quality Check:** The third level of quality control is removal of suspicious data by visual inspection.

2.1 Argo Real-time Quality Control (RTQC) Tests

Because of the requirement for delivering data to users within 24 hours of the float reaching the surface, the QC procedures on the real-time data are limited and automatic. The tests followed are listed in the table1. There are total 18 RTQC tests conducted on the Argo profile data. Details of the tests and the assignment of the quality flags are given in Appendix-1. More detail on the tests can also be found in IOC Manuals and Guides #22 and wong et al., (2012) . These real-time QC are done at INCOIS for floats deployed by India. Similarly, floats deployed by other countries will undergo the RTQC at their respective data centres and the data from these flots are archived at INCOIS on a day to day basis.

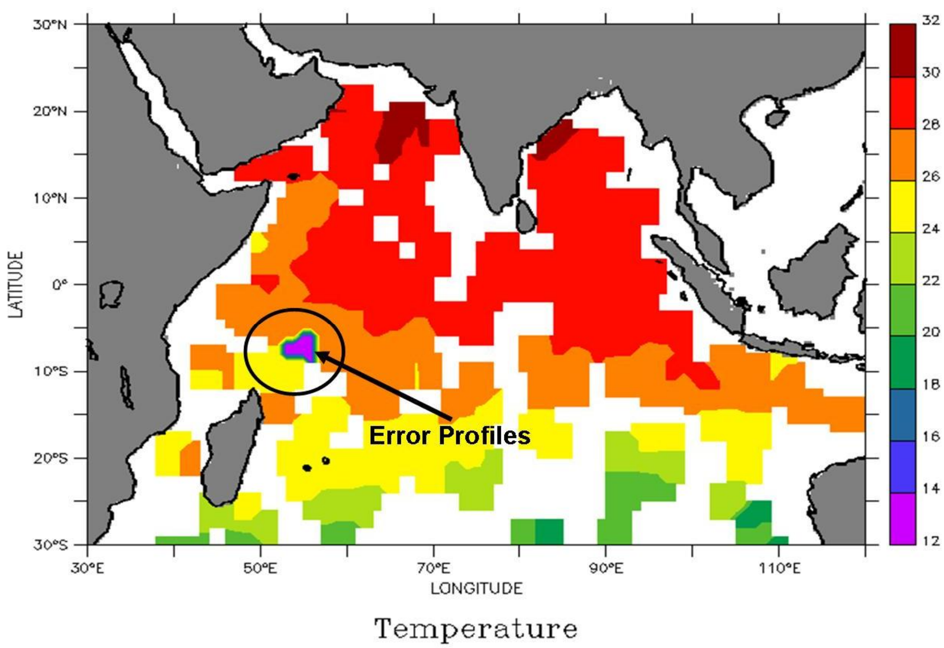

Figure2 Error profiles appearing as bull eye in the objectively analysed product. The error profiles are shown with in the circle pointed by arrow.

Even though majority of the profiles pass the real time QC test, there can be cases where in the profiles are incorrectly assigned flags (eg: salinity offsets). For example, incorrect metadata for the conversion from counts to physical values could be used which results in assignment of bad quality flags. Profiles from such a float could pass all the automatic tests, although the data are not correct. Also as most of the automatic QC tests rely on the records above and below for assigning a flag, a bad value in one record might cause record above and below to be assigned a bad flag.

2.2 Quality Control using Objective Analysis

To over come some of the problems with the automatic QC procedures, data is further subjected to additional quality checks. Objective analysis is used for identifying profiles with bad T/S values. All the available T/S profiles are objectively gridded. Kessler and McCreary (1993) method is used for objective analysis of the Argo T/S profiles on to regular grids. Details of this method is given in Appendix-2. Bad T/S profiles appear as bullôs eye like structure in the objectively analysed output. Figure 2 shows a typical objective analysis output where the bad T/S profiles are clearly shown as a bullôs eye (within the circle pointed by arrow). Once the bad profiles are pinpointed by the objective analysis, they can be visually checked for assigning correct quality flags using visual quality control tool.

2.3 Visual Quality Control

All the profiles which appear as bull eyes in the objective analysis product (described above) are isolated and subjected to further visual treatment. In this method, all the Argo T/S profiles that are identified to be appearing as bull's eye are visually checked for their quality using a visual quality control tool. The profiles under examination are overlaid on mean T/S profiles and associated standard deviations obtained world ocean atlas 2001 (WOA01) climatology (Conkright et al., 2002) .

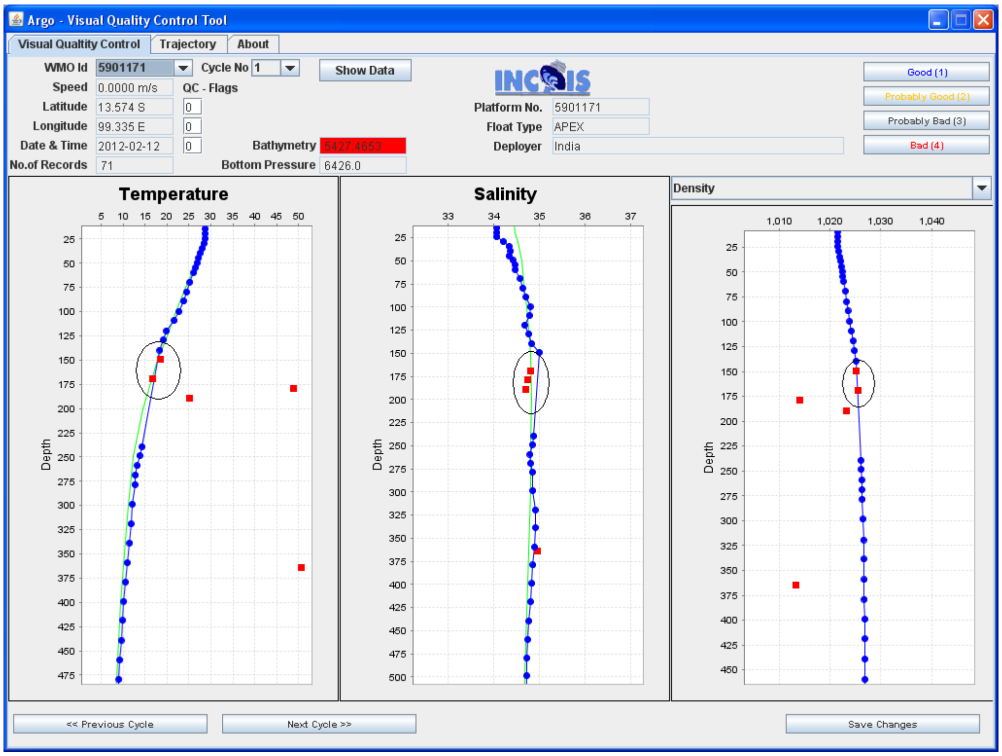

Figure.3: Typical example of profiles with flags assigned by real time QC. Blue dots are Argo observations and green thick lines are climatological means. Red dots are records under observation for their quality.

This visual quality control system was developed in house and is being used for quality control of CTD and XBT profiles. Udaya Bhaskar et al., (2012) describes the details of the VQC system. Argo profiles which are deviating beyond 2 standard deviations from the mean obtained from the WOA01 are assigned a bad flag. Figure 3 shows a typical case of T/S profiles which are checked visually for the correctness of the quality flags assigned. This way the profiles are thoroughly checked for their quality visually and flags are corrected. At the end of this exhaustive quality process, the good quality T/S profiles are utilized in generation of objectively analysed gridded products with a spatial resolution of 1° x 1° and temporal resolution of 10 days and month. Various value added products viz., Heat Content, Mixed Layer Depth, D20, Geostrophic Currents etc are derived using this gridded product. Details about gridding procedure and derived value added products can be obtained from Udaya Bhaskar et al., (2007). The objectively analysed gridded product for the period 2002 to 2012 is made available on INCOIS Live Access Server (www.las.incois.gov.in) in NetCDF format.

3. Discussion and summary

Quality control is a important part of ocean data assimilation system. As the automated QC of Argo temperature and salinity profiles are sometimes susceptible to errors, a system consisting of additional quality control procedures was developed and is put to extensive use at INCOIS. In this system all the profiles are passed through three mode of QC procedures viz., RTQC, objective analysis and visual quality control. Compared to the purely automated system, this systems requires a manual intervention (for visual inspection) preferably by experts in the field of oceanography hence the system is deemed as semi-automatic. Overall, it appears to be performing well although there is scope for improvement in any QC system. As a by product of this, objectively analysed gridded product on 10 day and monthly scale is being generated and made available on INCOIS Live Access Server in NetCDF format.

Acknowledgments

Autors are grateful to Dr. SSC Shenoi, Director, INCOIS for his encouragement and providing the facilities to carry out the work. Authors are also grateful to Dr Christine Coatanoan and Thierry Carval of Coriolis/IFREMER for their useful discussion on the Visual Quality Control. Argo data is made freely available by the Argo community.

References

Argo Science Team, 2001. The Global Array of Profiling Floats, in observing Oceans in 21st century. C.Z. Koblinsky and N.R. Smith, eds., Godae Proj. Off., *Bur. Meteorol*., Melbourne, Australia, (2001) 248 ï 258.

Bruce Ingleby and Matt Hudleston, 2007. Quality control of ocean temperature and salinity profiles - Historical and real-time data, Journal of Marine systems, 65, 158 - 175.

Conkright, M.E., R. A. Locarnini, H.E. Garcia, T.D. OöBrien, T.P. Boyer, C. Stephens, J.I. Antonov, 2002. *World Ocean Atlas 2001: Objective Analyses, Data Statistics, and Figures, CD-ROM Documentation.* National Oceanographic Data Center, Silver Spring, MD, 17 pp

Kessler and McCreary, 1993. The Annual Wind-driven Rossby Wave in the Subthermocline Equatorial Pacific, Journal of Physical Oceanography, Vol 23, 1192 - 1207.

Levitus, S., 1982. Climatological Atlas of the World Ocean. NOAA Professional Paper, Vol. 13. U.S. Government Printing Office, Washington, D.C.

Ravichandran M., P. N. Vinaychandran, S. Joseph and K. Radhakrishnan, 2004. Result from the first Argo float deployed by India, Current. Science, Vol 86, 651 ï 659.

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.

Udaya Bhaskar, T.V.S., E. Pattabhi Rama Rao, R. Venkat Shesu, R. Devender, 2012. GUI based interactive system for quality control of Argo data, Tech. Rept, INCOIS-DMG-TR-2012-02.

Wong, A., R. Keeley, T. Carval and the Argo Data Management Team. 2012. Argo quality control manual, Ver. 2.7, Report, 47 pp.

Appendix – 1

Automatic Real Time Quality Control procedures

1. Platform identification: Each float should have an unique valid identifier provided by World Meteorological Organization (WMO).

2. Impossible date/time test: Year must be greater than 1996; month in the range of 1 to 12; date must in the expected range for the month; hours in range 0 ï 23; minutes in the range 0 ï 59.

3. Impossible location test: The latitude (longitude) must be in the limits -90 to 90 (0 to 360).

4. Position on land test: The floats must be located in the ocean. ETOPO2 bottom topography is used for this test.

5. Impossible speed test: Surface and subsurface drift speeds must not exceed 3 m s^{-1}.

6. Global range test: Temperatures must be in the range of -2.5° to 40.0° C and salinity must be from 2 to 41 psu.

7. Regional range test: Temperatures from floats in the Red Sea (Mediterranean Sea) must range from 21.7° to 40.0° C (10.0° - 40.0° C) and salinity ranges must be from 2.0 to 41.0 (2.0 to 41.0 psu).

8. Pressure increasing test: The pressure must increase monotonically.

9. Spike test: $|V_2-(V_3+V_1)/2| - |(V_3-V_1)/2|$ for a value V_2, where V_1 and V_3 are the values above and below V_2, which may not exceed prescribed limits. Above 500 dbar, the limit for temperature (salinity) is 6°C (0.9) and below 500 dbar the limits are 2°C (0.3).

10. Top and bottom spike test: This test is obsolete now.

11. Gradient test: The test value $|V_2 - (V_3+V_1)/2|$ for a value V_2 may not exceed prescribed limits. Above 500 dbar, the limit for temperature (salinity) is 9.0°C (1.5) and below 500 dbar the limits are 6.0°C (0.5).

12. Digit rollover test: A specific number of bits are allocated for the storage of temperature and salinity values in a float. When the number is exceeded, stored values rollover to the lower end of the range. This rollover when detected is compensated for in the processing algorithm.

13. Stuck value test: This test checks for constant temperature or salinity values throughout the profile.

14. Density inversion : This test computes the density at all pressure levels from the observed temperature and salinity values and tests for hydrostatic stability.

15. Grey list: A list generated based on the history of a float. When a float sensor has systematic problems it is placed on this list.

16. Gross salinity or temperature sensor drift: If the average temperature (salinity) from the last 100 dbar of two adjacent profiles exceeds 1°C (0.5), then the profile is considered to be bad.

17. Frozen profile test: If floats produce five consecutive profiles with very small differences throughout the entire water column (i.e., of the order of 0.001 for salinity and of the order of 0.01°C for temperature) they are candidates for the gray list.

18. Deepest pressure test : This test requires that the profile has pressures that are not higher than DEEPEST_PRESSURE plus 10%. DEEPEST_PRESSURE value comes from the meta-data file of the float.

Appendix - 2

Kessler and McCreary objective analysis method

This objective analysis method was used for eliminating spurious T/S profiles which appear as bull eye like structure in the gridded product. The objective analysis was carried out in two steps. First the temperature data T_n (x_n, y_n, z_n, t_n) from each profile n were linearly interpolated to standard depths (1 m from surface to 2000 m) there by creating a modified data set $T^{\hat{o}}_n$ (x_n, y_n, z_n, t_n). This interpolation was done only when two sample in a profile are within a selected vertical distances which increased from 5 m in the surface to 100 between 500 ï 2000 m. Second, in a separate computation at each Levitus standard depth Z_0 (0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000 m) the temperature $T^{\hat{o}}_n$ were mapped from irregular grid locations (x_n, y_n, z_n) to regular grid (x_0, y_0, z_0) locations with a grid spacing of 1º X 1º . Specifically the value of the gridded temperature $T^{\hat{o}\hat{o}}$ at each grid point (x_0, y_0, t_0) was estimated by the operation

$$T''(x_0, y_0, t_0) = \frac{\sum_{n=1}^{N_p} T'_n W_n}{\sum_{n=1}^{N_p} W_n} \qquad (1)$$

where N_p is the total number of profiles within the influence region of a particular grid point. The Guassian weight function W_n is given by

$$W_n(x_n, y_n, t_n) = exp\left\{-\left[\left(\frac{x_n - x_0}{X}\right)^2 + \left(\frac{y_n - y_0}{Y}\right)^2 + \left(\frac{t_n - t_0}{\tau}\right)^2\right]\right\} \quad (2)$$

As pointed by Kessler and McCreary (1993) this operation is similar to a single iteration of objective mapping as used by Levitus (1982). Visualizing the three dimensional grid (x_0, y_0, z_0) with data points $T^{\hat{o}}_n(x_n, y_n, z_n)$ scattered irregularly through it, the mapping operation appears as a ellipsoid moving from grid point to grid point averaging the points that fall within that ellipse. Each data point falls within summation of several grid points, weighted according to the distance. In the regions of very sparse sampling, a single data point may be the only information for one or several grid points. If no data points fell within the ellipsoid at a grid point, then that was left blank.