ICES Guidelines for Profiling Float Data

(Compiled January 2001; revised April 2006)

Profiling floats are neutrally buoyant devices that periodically surface to transmit data to a satellite system. The float drifts along a specified pressure surface for some period of time (typically seven to 10 days) after which it surfaces. During the ascent, sensors attached to the float measure ocean parameters such as temperature and conductivity. While at the surface, the float transmits the data, along with current float position and time information, to a satellite system. After a repeated transmission over a period of up to one day, the float descends to it predetermined pressure level (called also "parking depth") and repeats the cycle.

Floats were developed during the World Ocean Circulation Experiment (WOCE). Present day floats can obtain depths of 2000 m, and complete about 100 cycles. Accuracy of about 0.01°C in temperature and 0.01 in salinity are obtainable.
Profiling floats are now (2006) mainly deployed as part of the ARGO project and data collected in this project must be supplied to one of the two ARGO global data centres (Coriolis and US-GODAE servers)

1.0   RECEIVING DATA

The Data Centres require the following information to be supplied by the data supplier together with the data. When receiving data, the Data Centres of the ICES community shall strive to meet the following guidelines.

Note that the following guidelines address two distinct data types collected by profiling floats: profile data and surface trajectory (Lagrangian) data. As well, note that the Data Centres may receive these data by several mechanisms: (I) raw data from the transmission service (e.g. Service Argos, ORBCOMM), (II) real-time profile messages via the Global Telecommunication System (GTS) and (III) delayed-mode profile data from the Principal Investigators (PI).

1.1   Data Standard

Quality controlled float data provided by Principal Investigators to the Data Centres should contain:
- A full description of the format used.
- Metadata information about the calibration (equations and coefficients) - see, the detailed description in Argo Data Management User's Manual at http://www.coriolis.eu.org/cdc/argo_rfc.htm - and processing techniques used for each parameter
- Quality controlled profile data, reported at the original sampling interval of the profiling float
- Quality controlled trajectory data
- Description of the quality control procedures applied

The trajectory data may have been picked up by the Data Centre from the transmission service (Argos, ORBCOMM, etc.). These data pass through automatic quality control procedures (including around 15 individual tests) at the Data Centre, checking for position errors by examining computed speeds. The delayed mode trajectory data is passed to the PI for additional quality control and returned to the Data Centre after the PI quality control is complete.

It is recommended that the profile data are managed by the Data Centre in a continuously managed database.   An overview of the data management practices for delayed mode profile data in a continuously managed database is provided in Annex A (see Wilson, 1998).

All parameters must be clearly specified and described. If parameter codes are to be used, then the source data dictionary must be specified in the metadata documentation. Parameter units must be clearly stated. Parameter scales must be noted where applicable. If computed values are included, the equations used in the computations should be stated.

All relevant calibrations should be applied to the delayed mode profile data including laboratory and field calibrations. Instrument calibration data should be included in the data file. The data should be fully checked for quality and flagged for erroneous values such as spikes, gaps, etc. An explicit statement should be made of the checks and edits applied to the data. If any data values have been removed, the levels of the removed values should be noted. Particular attention should be paid to the calibration and performance of the conductivity cell (see Bacon et al., 1998).

Sufficient self-explanatory information and documentation should accompany the delayed mode profile data so that they are adequately qualified and can be used with confidence by scientists/engineers other than those responsible for its original collection, processing and quality control.

The PI/data supplier should ensure that the following be provided with the delayed mode profile data submission:
- All data values should be expressed in oceanographic terms, in SI units, which should be clearly stated
- Time reported in UTC is strongly recommended
- The units used for the measured parameters should be clearly described and consistent

1.2   Format Description

Real-time profile message transmitted over the Global Telecommunications System (GTS) may be coded in either TESAC (ASCII) or BUFR (binary) format. The coded message will typically consist of all points in the profile.

The data formats for the exchange of profile data may initially vary. Data can be supplied in a fully documented ASCII format. Argo data are supplied in netCDF format (example of Argo files can be downloaded at ftp://ftp.ifremer.fr/ifremer/argo/dac/coriolis/)
.

Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is strongly recommended. If netCDF is the delivery format, then one profile can be contained in a single netCDF file. A metadata file (ASCII) should accompany the data (Argo data management user's manual at http://www.coriolis.eu.org/cdc/argo_rfc.htm). In addition, the contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange (http://ocean.ices.dk/formats/GETADE_Guidelines.aspx) prepared by the IOC's Group of Experts on the Technical Aspects of Data Exchange (GETADE) and available from RNODC Formats.

1.3   Collection Details

Data transfer from PIs to the Data Centres consists of a one-time pre-deployment metadata transfer and an operational data transfer of delayed mode data.

Before deployment of a profiling float, the PI or data supplier needs to inform the Data Centre of the intended deployment. Other pertinent metadata will also be required by the Data Centre including:

- Deployment platform name
- Country, organisation, PI
- Project name
- Anticipated deployment date and general location
- Float number
- WMO number
- Transmission ID number
- Transmission service provider
- Nominal parking depth
- Surfacing interval
- Sensor resolutions
- Vertical levels
- Metadata information about the instruments and sensors (type and manufacturer, serial and model numbers, board type and serial number, software version)
- Metadata information about the data precision and final accuracy

Any additional information of use to secondary users which may have affected the data or have a bearing on its subsequent use.

2.0   VALUE ADDED SERVICE

When processing and quality controlling data, the Data Centres of the ICES community shall strive to meet the following guidelines.

2.1   Quality Control

Details on the quality control of profile data are described in IOC MANUALS AND GUIDES #22 (http://www.meds-sdmm.dfo-mpo.gc.ca/ALPHAPRO/gtspp/qcmans/MG22/guide22_e.htm). Although the Manual is intended for the quality control of real-time profiles, the same principles apply to delayed mode data. Along with other relevant information, the Guide provides specific details on quality flagging, quality control tests, duplicates management, implementation details, as well as additional references on the treatment of profile data by other data centres.

There are three main components to the quality control of float profile data. All three components are used at the ICES data centres to quality control the float profile datasets.

The first component examines the characteristics of the float track looking to identify errors in either position or time.

The second component examines the various profiles of observations to identify values that appear to be in error. Knowledge of the different types of real and erroneous features is critical. This knowledge, when combined with a local knowledge of water mass structure, statistics of data anomalies, thermocline characteristics, and cross validation with climatological data, ensures a data set of the best possible quality.

The third component is software to identify duplicate profiles. Duplicate profiles occur either by having received the data more than once, or because real-time messages arrive before the delayed-mode data on which the real-time message was based.

Quality control findings for the original data set are shared with the data originator to maintain consistency and uniqueness of the mutual data set and improve its overall quality. Complete description of the ARGO quality control procedure are available at: ARGO Quality Control

2.2   Problem Resolution

The quality control procedures followed by the Data Centres will typically identify problems with the data and/or metadata. The Data Centre will resolve these problems through consultation with the originating PI or data supplier. Other experts in the field or other Data Centres may also be consulted.

2.3   History Documentation

All quality control procedures applied to a dataset are fully documented by the Data Centre. As well, all quality control applied to a dataset should accompany that dataset. All problems and resulting resolutions will also be documented with the aim to help all parties involved; the Collectors, Data Centre, and Users. A history record will be produced detailing any data changes (including dates of the changes) that the Data Centre may make.

3.0   PROVIDING DATA AND INFORMATION PRODUCTS

When addressing a request for information and/or data from the User Community, the Data Centres of the ICES community shall strive to provide well-defined data and products. To meet this objective, the Data Centres will follow these guidelines.

3.1   Data Description

The Data Centre shall aim to provide well-defined data or products to its clients. If digital data are provided, the Data Centre will provide sufficient self-explanatory information and documentation to accompany the data so that they are adequately qualified and can be used with confidence by scientists/engineers other than those responsible for their original collection, processing and quality control.
- A data format description fully detailing the format in which the data will be supplied
- Parameter and unit definitions, and scales of reference
- Definition of flagging scheme, if flags are used
- Relevant information included in the metadata or data file (e.g. ship, cruise, project, start and end date, nominal parking pressure, surfacing interval, etc.)
- Data history document (as described in 3.2 below)

3.2   Data History

A data history document will be supplied with the data to include the following:
- A description of data collection and processing procedures as supplied by the data collector (as specified in Section 1.1 and 1.3)
- Quality control procedures used to check the data (as specified in Section 2.1)
- Any problems encountered with the data and their resolution and modification date
- Any changes made to the data and dates of these changes

Any additional information of use to secondary users which may have affected the data or have a bearing on its subsequent use should also be included.

The Argo format includes all the historical information about the data.

3.3   Referral Service

ICES member research and operational data centres produce a variety of data analysis products and referral services. By dividing ocean areas into regions of responsibility, and by developing mutually agreed guidelines on the format, data quality and content of the products, better coverage is obtained. By having the scientific experts work in ocean areas with which they are familiar, the necessary local knowledge finds its way into the products. Data and information

products are disseminated as widely as possible and via a number of media including mail, electronic mail and bulletin boards.

If the Data Centre is unable to fulfil the client's needs, it will endeavour to provide the client with the name of an organisation and/or person who may be able to assist. In particular, assistance from the network of Data Centres within the ICES Community will be sought.

REFERENCES

Argo Data Management Handbook Version 1.2,
Argo User's Manual , Version 2.1
Argo Quality Control Manual, Version 2.
available at :http://www.coriolis.eu.org/cdc/argo_rfc.htm

Bacon, S., L. Centurioni and W. J. Gould, 1998: Evaluation of Profiling ALACE Float Performance. Southampton Oceanography Centre Internal Document, No. 39, 72 pp. (www.soc.soton.ac.uk/JRD/HYDRO/shb/posters/BCG.PALACE.Report.pdf)

GTSPP real time quality control manual. 1990. <u>IOC MANUALS AND GUIDES Number 22. (http://www.meds-sdmm.dfo-mpo.gc.ca/ALPHAPRO/gtspp/qcmans/MG22/guide22_e.htm)</u>

Rual, Pierre. 1989. For a Better XBT Bathy-message: Onboard Quality Control, plus a New Data Reduction Method, Proceedings of the Western Pacific International Meeting and Workshop on TOGA COARE, Edited by Joel Picaut, Roger Lukas and Thierry Delcroix, Centre ORSTIM de Noumea, New Caledonia, May 24-30, 1989.

Wilson, J.R. 1998. Global Temperature-Salinity Profile Programme (GTSPP) - overview and future, Intergovernmental Oceanographic Commission technical series 49, SC-98/WS/59.

ANNEX A

Both the real-time messages and delayed mode data are available at the sample interval set by the manufacturer. This interval is typically between 4 and 100 dbars depending on location in the water column. The delayed mode data undergo maximum calibration and quality control often incorporating site specific knowledge and experience of the PI. Real-time messages are those profile data that have undergone only those automatic, bulk quality control tests implemented by the Data Centre. The extensive quality control incorporating site specific knowledge and experience of the PI. Real-time messages are most useful to those involved in operational forecasts, while delayed mode data are more useful to scientists. For float profile data it is ideally about five months before the delayed mode data are submitted to data centres.

To manage profile data, a Continuously Managed Database (CMD) system is implemented. As data are acquired in both real-time and delayed mode they are added to the database. Calibrated and quality controlled delayed mode data replaces the message obtained in near real-time. The CMD therefore holds the most current and highest quality data set at all times. The database is continuously refined as additional quality checks are undertaken. Observations that have passed quality control and entered the database are not removed but are flagged to indicate that a higher quality version of the observation exists in the database.