6Th Framework of EC DG Research

SeaDataNet

DATA QUALITY CONTROL PROCEDURES

Version 2.0
May 2010

Date: 14 May 2010
Website: **www.seadatanet.org**

Quality Control Standards for SEADATANET

Contents

1. Introduction
2. Why is quality control is needed?
3. Information to accompany data
   - 3.1 Metadata
   - 3.2 Parameter vocabularies
4. Automatic checks
5. “Scientific” quality control
   - 5.1 CTD (temperature and salinity)
   - 5.2 Currents
   - 5.3 Wave data
   - 5.4 Sea Level
   - 5.5 Chemical sample data (nutrients, oxygen)
   - 5.6 Biological data, etc.,
6. Quality control flags
7. Documentation
8. References

Appendix 1: Examples of ICES Data Guidelines

1. Introduction

SeaDataNet is a European Infrastructure (DG-Research – FP6) project which is developing and operating a Pan-European infrastructure for managing, indexing and providing access to ocean and marine environmental data sets and data products (e.g. physical, chemical, geological, and biological properties) and for safeguarding the long term archival and stewardship of these data sets. Data are derived from many different sensors installed on research vessels, satellites and in-situ platforms that are part of various ocean and marine observing systems. Data resources are quality controlled and managed at distributed data centres that are interconnected by the SeaDataNet infrastructure and accessible for users through an integrated portal. The data centres are mostly National Oceanographic Data Centres (NODCs) which are part of major marine research institutes that are developing /operating national marine data networks, and international organizations such as IOC/IODE and ICES. The data sets managed come from various sources and time periods. This imposes strong requirements towards ensuring quality, elimination of duplicate data and overall coherence of the integrated data set. This is achieved in SeaDataNet by establishing and maintaining accurate metadata directories and data access services, as well as common standards for vocabularies, metadata formats, data formats, quality control methods and quality flags.

The Earth's natural systems are complex environments in which research is difficult in most instances and where many natural factors and events need to be taken into consideration. Especially complex are the aquatic environments which have specific research obstacles to overcome, namely deep, dark and often turbulent conditions. Good quality research depends on good quality data and good quality data depends on good quality control methods. Data can be considered 'trustworthy' after thorough processing methods have been carried out. At this stage they can be incorporated into databases or distributed to users via national or international exchange.

Data quality control essentially and simply has the following objective:

"To ensure the data consistency within a single data set and within a collection of data sets and to ensure that the quality and errors of the data are apparent to the user who has sufficient information to assess its suitability for a task."
(IOC/CEC Manual, 1993)

If done well, quality control brings about a number of key advantages:

- <u>Maintaining Common Standards</u>
  There is a minimum level to which all oceanographic data should be quality controlled. There is little point banking data just because they have been collected; the data must be qualified by additional information concerning methods of measurement and subsequent data processing to be of use to potential users. Standards need to be imposed on the quality and long-term value of the data that are accepted (Rickards, 1989). If there are guidelines available to this end, the end result is that data are at least maintained to this degree, keeping common standards to a higher level.

- <u>Acquiring Consistency</u>
  Data within data centres should be as consistent to each other as possible. This makes the data more accessible to the external user. Searches for data sets are more successful as users are able to identify the specific data they require quickly, even if the origins of the data are very different on a national or even international level.

- <u>Ensuring Reliability</u>
  Data centres, like other organisations, build reputations based on the quality of the services they provide. To serve a purpose to the research community and others their data must be reliable, and this can be better achieved if the data have been quality controlled to a 'universal' standard. Many national and international programmes or projects carry out investigations across a broad field of marine science which require complex information on the marine environment. Many large-scale projects are also carried out under commercial control such as those involved with oil and gas and fishing industries. Significant decisions are made, and theories formed, on the assumption that data are reliable and compatible, even when they come from many different sources.

2. Why is quality control is needed?

It is beneficial to publish the current good practice and distribute the information widely in order that a more standardized approach can be realised. The procedures include a variety of automatic tests which are often carried out in real time and also a more scientific quality control, checking for unexpected anomalies in the time series or profile, or in derived parameters. Quality control extends beyond these procedures mentioned to include the documentation of the data sets.

Quality control is also related to issues such as the availability of data in real-time. If data are inspected every day or, in advanced systems, if data can be flagged for errors by automatic software, then faults can be rapidly attended to and fixed. This contrasts with the more traditional form of carrying out all of the procedures in delayed mode, where errors are detected a considerable time after they occur.

This manual draws on existing documents including those produced by international organisations (e.g. IOC's International Oceanographic Data and Information Exchange (IODE) programme, JCOMM Data Management Programme Area (DMPA) and the International Council for the Exploration of the Sea's Working Group on Data and Information Management (ICES WGDIM)), international projects and programmes (e.g. WOCE, JGOFS, GTSPP, GOSUD, Argo, GLOSS, etc.), other European projects (in particular MyOcean for real time quality control) national programmes and expertise from national oceanographic data centres and marine research organisations to derive a set of recommended standards for quality control of a variety of marine data. This will result in data sets which have been acquired and processed to agreed standards and will allow future researchers to better define confidence limits when applying these data. These should be reviewed and updated at regular intervals.

3. Information to accompany data

3.1 Metadata

Data are submitted to marine data centres for the purpose of long term viability and future access. This requires the data set to be accompanied by key data set information (metadata). Detailed metadata collation guidelines for specific types of data are either available or under development to assist those involved in the collection, processing, quality control and exchange of those data types.

A summary checklist is provided below. For all types of data the following information is required:

- **Where** the data were collected: location (preferably as latitude and longitude) and depth/height
- **When** the data were collected (date and time in UTC or clearly specified local time zone)
- **How** the data were collected (e.g. sampling methods, instrument types, analytical techniques)
- **How** you refer to the data (e.g. station numbers, cast numbers)
- **Who** collected the data, including name and institution of the data originator(s) and the principal investigator
- **What** has been done to the data (e.g. details of processing and calibrations applied, algorithms used to compute derived parameters)
- **Watch points** for other users of the data (e.g. problems encountered and comments on data quality)

The ICES Working Group on Data and Information Management (WGDIM) has developed a number of data type guidelines which itemize these elements that are required for thirteen different data types (see table below). These Data Type Guidelines have been developed using the expertise of the oceanographic data centres of ICES Member Countries. They have been designed to describe the elements of data and metadata important to the ocean research community. These guidelines are targeted toward physical-chemical-biological data types collected on oceanographic research vessel cruises. Each guideline addresses the data and metadata requirements of a specific data type. This covers three main areas:

- What the data collector should provide to the data centre (e.g. collection information, processing, etc)
- How the data centre handles data supplied (e.g. value added, quality control, etc)
- What the data centre can provide in terms of data, referral services and expertise back to the data collector.

A selection of these guidelines, in particular for those data types that are not yet dealt with in detail here, are included in Appendix 1 of this document.

ICES Data Type Guidelines		
CTD	Moored ADCP	Moored Current Meter
Shipborne ADCP	Seasoar (Batfish)	Surface (Underway)
Water Level	XBT	Net Tow (Plankton)
Surface Drifting Buoy	Profiling Float and Drifting Buoy	Discrete water sample
Multibeam echosounder data		

3.2 Parameter Usage Vocabulary

SeaDataNet has adopted and built upon the BODC Parameter Usage Vocabulary (formerly the BODC Parameter Dictionary). Elements are used for labelling data as they are submitted to a data centre or stored within a research institute. Instead of using non-standard descriptions for parameters, individual codes are assigned from the dictionary, and standardisation is achieved. This includes information about what was measured and can include additional information such as how the measurement was made.

During the 1990s BODC was heavily involved in the Joint Global Ocean Flux Study (JGOFS), which required rapid expansion of the vocabulary to about 9000 parameters. When BODC first started managing oceanographic data in the 1980s, fewer than twenty parameters were handled. This rapid increase in the number of parameters forced BODC to adopt a new approach to parameter management and develop the vocabulary. It now comprises entries for more than 21,000 physical, chemical, biological and geological parameters. Sometimes a single water bottle sample has been analysed for several hundred parameters. The BODC Parameter Usage Vocabulary (P011) is freely available from the NERC vocabulary server (www.bodc.ac.uk/products/web_services/vocab/), through web services and via the SeaDataNet web site (seadatanet.maris2.nl/v_bodc_vocab/welcome.aspx).

The NERC Vocabulary Server provides access to lists of standardised terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community.

Using standardised sets of terms (otherwise known as "controlled vocabularies") in metadata and to label data solves the problem of ambiguities associated with data markup and also enables records to be interpreted by computers. This opens up data sets to a whole world of possibilities for computer aided manipulation, distribution and long term reuse.

An example of how computers may benefit from the use of controlled vocabularies is in the summing of values taken from different data sets. For instance, one data set may have a column labelled "Temperature of the water column" and another might have "water temperature" or even "temperature". To the human eye, the similarity is obvious but a computer would not be able to interpret these as the same thing unless all the possible options were hard coded into its software. If data are marked up with the same terms, this problem is resolved.

In the real world, it is not always possible or agreeable for data providers to use the same terms. In such cases, controlled vocabularies can be used as a medium to which data centres can map their equivalent terms.

The controlled vocabularies delivered by the NERC Vocabulary Server contain the following information for each term:

1) Key — a compact permanent identifier for the term, designed for computer storage rather than human readability
2) Term — the text string representing the term in human-readable form
3) Abbreviation — a concise text string representing the term in human-readable form where space is limited
4) Definition — a full description of what is meant by the term

All of the vocabularies are fully versioned and a permanent record is kept of all changes made.

4. Automatic checks

A number of basic automatic checks should be carried out on all data. These include date and time, position, and range checks. The MyOcean in situ Thematic Assembly Centre has produced a suite of documents for real time quality control covering temperature and salinity, currents (moored and drifters), sea level and biochemical (chlorophyll_a fluorescence, oxygen and nutrient measurements) data. In addition, GTSPP has recently (September 2009) revised its real time quality control manual which lays out in detail the automatic tests to be carried out on temperature and salinity data.

(i) Date and time of an observation has to be valid:

- Year 4 digits – this can be tuned according to the data
- Month between 1 and 12
- Day in range expected for month
- Hour between 0 and 23
- Minute between 0 and 59

(ii) Latitude and longitude have to be valid:

- Latitude in range -90 to 90
- Longitude in range -180 to 180

(iii) Position must not be on land

- Observation latitude and longitude located in ocean
- The test requires that the observation latitude and longitude from the profile measurement be located in an ocean. Use can be made of any file that allows an automatic test to see if data are located on land. We suggest use of at least the 2-minute bathymetry file that is generally available. This is commonly called and can be downloaded from http://www.ngdc.noaa.gov/mgg/global/etopo2.html.

(iv) Global range test

- Tests that observed parameter values are within the expected extremes encountered in the oceans

(v) Regional range test

- Tests that observed parameter values are within the expected extremes encountered in particular regions

(vi) Deepest pressure

- Tests that profile does not contain pressures higher than the highest value expected

5. "Scientific" quality control

Further quality control is carried out on the data sets, and may be dependent on the data type. There is often a subjective element in this process. This type of quality control is described below for a number of data types and further information is given in Appendix 1 for a number of further data types.

5.1 CTD (temperature and salinity)

Much documentation exists for the quality control of temperature and salinity data. These data may come from a variety of sources including water bottles, CTDs, profiling floats and instruments attached to marine mammals.

For example, the ICES Guideline for CTD data provides a range of checks are carried out on the data by a data Centre to ensure that they have been imported into the Data Centre's format correctly and without any loss of information. For CTD data, these should include:

- Check header details (vessel, cruise number, station numbers, date/time, latitude/longitude (start and end), instrument number and type, station depth, cast (up or down)), data type/no. of data points)
- Plot station positions to check not on land
- Check ship speed between stations to look for incorrect position or date/time
- Automatic range checking of each parameter
- Check units of parameters supplied
- Check pressure increasing
- Check no data points below bottom depth
- Check depths against echo sounder
- Plot profiles (individually, in groups, etc)
- Check for spikes
- Check for vertical stability/inversions
- Plot temperature vs. salinity
- Check profiles vs. climatology for the region
- Check calibration information available

As already mentioned GTSPP has an extensive manual (IOC Manuals and Guides 22) documenting the quality tests to carry out on temperature and salinity data. Similarly, the Argo project has documented a set of real time quality control tests and also delayed mode quality control methods. GOSUD has also documented tests to be carried out on surface underway data. Some of these tests are those already described above; others are given below.

Example 1: Additional Argo Real time QC tests:

- Deepest pressure — Tests that the profile does not contain pressures higher than the highest value expected for a float.
- Pressure increasing — Tests that pressures from the profile are monotonically increasing.
- Spike — Tests salinity and temperature data for large differences between adjacent values.
- Gradient — Tests to see if the gradient between vertically adjacent salinity and temperature measurements are too steep.
- Digit rollover — Tests whether the temperature and salinity values exceed a floats storage capacity.
- Stuck value — Tests for all salinity or all temperature values in a profile being the same.
- Density inversion — Tests for the case where calculated density at a higher pressure in a profile is less than the calculated density at an adjacent lower pressure.
- Grey list — Tests whether the sensor identifier is present in a list that has been collated to identify sensors which are experiencing problems.
- Sensor drift — Tests temperature and salinity profile values for a sudden and important sensor drift.
- Frozen profile — Tests for the case where a float repeatedly produces the same temperature or salinity profile (with very small deviations).

Example 2: Additional GOSUD QC tests:
Test 8 : spike test
Differences between sequential measurements, where one measurement is quite different than adjacent ones, is a spike in both size and gradient.
Test value = | V2 - (V3 + V1)/2 | - | (V3 - V1) / 2 |
where V2 is the measurement being tested as a spike, and V1 and V3 are the values previous and next.

- Temperature: The V2 value is flagged when the test value exceeds 6.0 degree C.
- Salinity: The V2 value is flagged when the test value exceeds 0.9 PSU

Values that fail the spike test should be flagged as wrong and should not be distributed.

Test 9 : gradient
This test is failed when the difference between adjacent measurements is too steep.
Test value = | V2 - (V3 + V1)/2 |
where V2 is the measurement being tested as a spike, and V1 and V3 are the previous and next values.

- Temperature: The V2 value is flagged when the test value exceeds 9.0 degree C.
- Salinity: The V2 value is flagged when the test value exceeds 1.5 PSU

Values that fail the test (i.e. value V2) should be flagged as wrong.

Test 10 : climatology
Each measurement is compared to a climatology.
The test fails if | V1 – V2 | > 3 * Sigma
Where V1 : value to be controlled
V2 : value of the climatology
Sigma : standard deviation of the climatology
The climatology is Levitus, 1998, 1°x1°, monthly.

If the test fails, the data is flagged as "out of statistics" (flag 2). However, the data can be distributed.

Test 11 : instrument comparison

If two different sensors measure a same parameter, the difference between 2 measurements should not be greater than a fixed limit.

Example : on research vessels the difference between the temperature of the tank of the TSG and the measurement of the hull monted temperature sensor should be less than 1° Celsius. If the test fails, the measurements of both sensors are flagged as wrong.

5.2. Current meter data (including ADCP)

Screening Procedure

BODC’s in-house software for quality controlling current meter data comprises a visualisation tool called SERPLO (SERies PLOtting), developed in response to the needs of BODC whose mandate involved the rapid inspection and non-destructive editing of large volumes of data. SERPLO allows the user to select specific data sets and view them in various forms, to visually asses their quality. Displays include timeseries, depth series, a scatter plot for current meter data, an X-Y plot and a year’s display for tidal data. There is also a world map covering the locations of series. Screening essentially allows the quality control of data that we receive with checks being made to ensure that the data are free from instrument-generated spikes, gaps, spurious data at the start and end of the record and other irregularities, for example long strings of constant values. These problems are not immediately obvious when just looking at large columns of data. When suspicious values are seen, flags are applied to the data points in question as a warning to end users. BODC uses two types of flag, M and N. The M flag is assigned to suspicious values whereas the N flag is assigned to those values that are null. It is necessary to emphasise that these flags do not change the data; they purely highlight potential problems with the data, allowing the end user to decide the usefulness of the data. Screening is after all, a procedure based very much on instinct and perception and opinions will inevitably differ from person to person.

Time Series

The most useful screening-view for current meter data is the time series; a plot of the parameters measured over the time of the record. This is useful as the user can get an idea very quickly about whether the data looks reasonable or not judging by the average values of the parameter measured and the overall ‘noisiness’ of the plot. Using the time series all parameters can be visually ‘screened’ with the aim to look for anomalous values. Anomalous values are those which are out of character with the rest of the series and therefore unlikely to be a true representation. The most common are found as ‘spikes’, usually caused by a problem with the instrument as opposed to a sudden rapid change in the water conditions. ‘Spikes’ are usually singular points which are completely out-of-range when compared to the immediate surrounding values. It is possible that when there are a few data points within a single ‘spike’ the values may represent a true event and as such these points making up the spike are not generally flagged unless they are hugely out of character with the rest of the series. The figures below show examples:

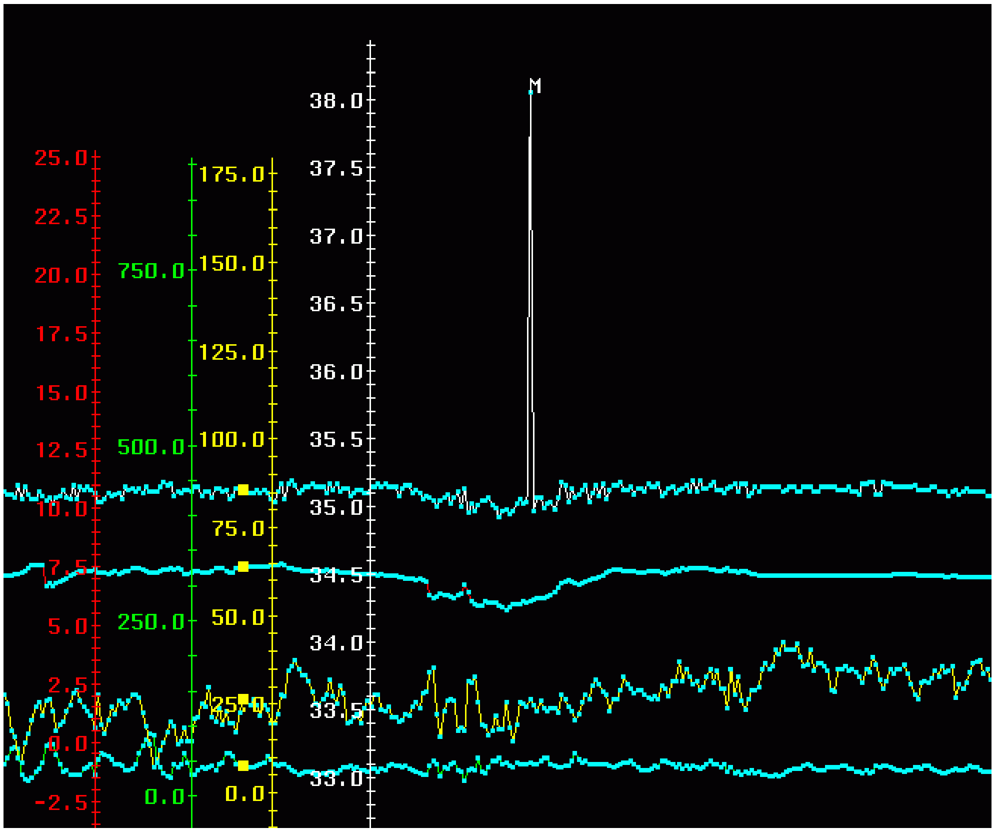

Figure 1 – Time Series plot showing a 'spike' in salinity and its appropriate flag

Figure 1 gives an example of a 'spike' in salinity within a series. The spike has been flagged as suspect with the customary 'M' flag. As can be seen only one point is out-of-range indicating it is likely to be an instrument error.

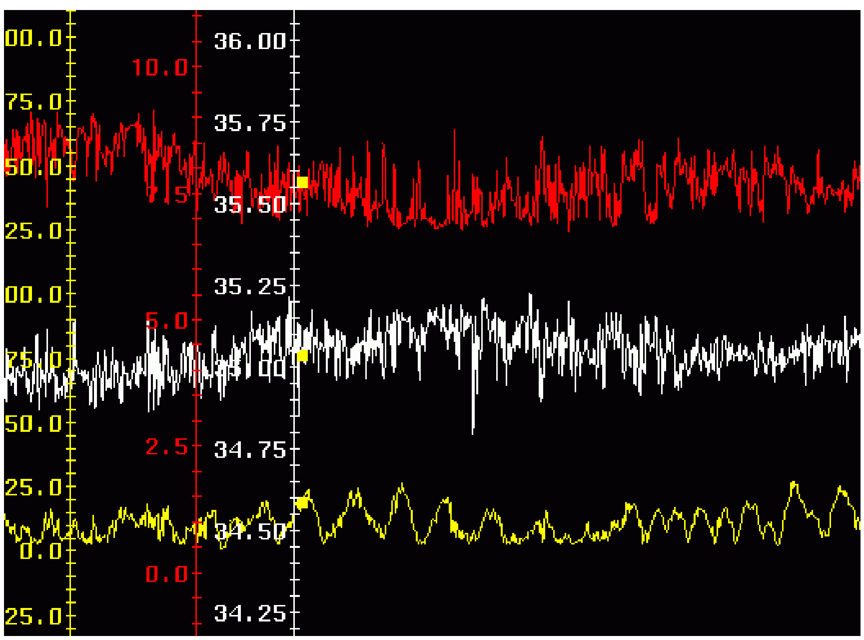

Figure 2 – Time series plot showing 'noisy' temperature, salinity and speed data

Figure 2 is an example of a situation where it is difficult to identify spikes because the surrounding data are noisy indicating a natural phenomenon having an effect on the localized water conditions.

It is helpful to screen related parameters, such as current speed and current direction, together to identify spurious values. It is common for related changes to occur with related parameters. If this is not seen it is often an indication that the event is not genuine and should be marked accordingly. For instance salinity is calculated from temperature. If there is a sudden change in salinity which is not seen in temperature it could just be due to a calculation error.

The derived velocity components can also be screened and indeed it is useful to check any suspect values in the speed or direction against the associated point in the u- and v components.

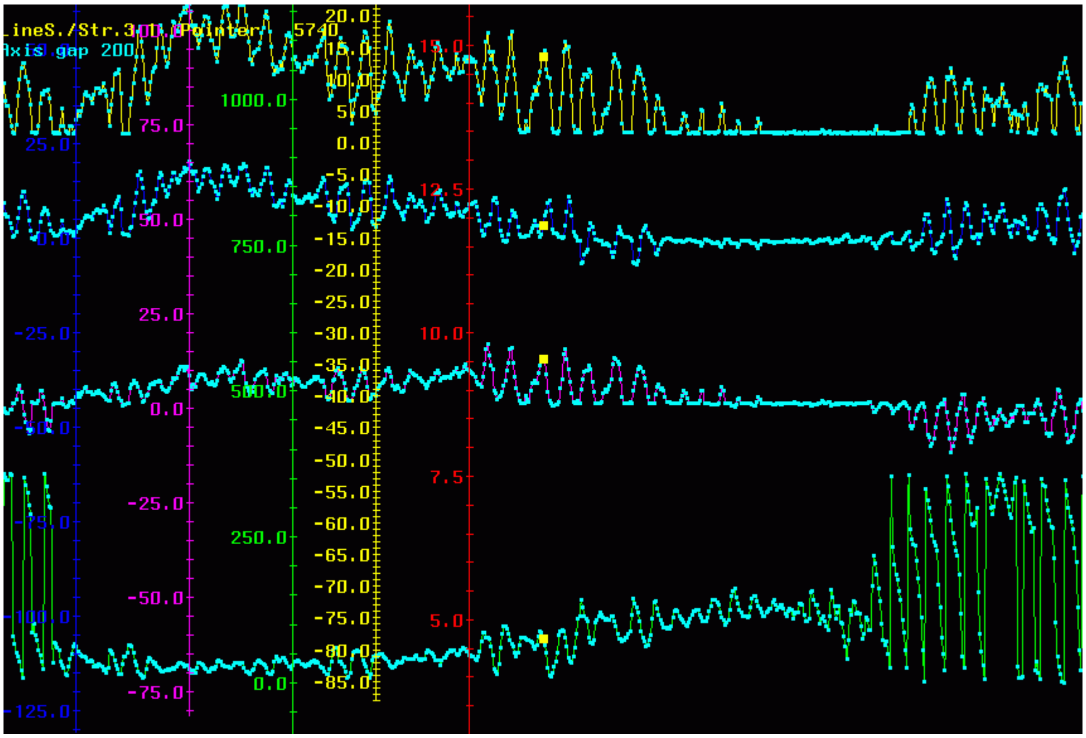

Figure 3 - Example of a time series with a possible rotor problem

(Legend: Current Speed (ms-1), North velocity component (ms-1), East velocity component (ms-1), Current Direction (°), Temperature (°C) (not shown))

Figure 3 shows a typical time series plot of current speed and direction. In this specific example the current speeds are at the meter threshold level throughout the series, as can be seen by the period near the end of the plot where current speed and the velocity components 'flatten out' to read near zero. This could mean that the rotor is being hindered in some way, therefore not turning freely and this should be noted in subsequent documentation.

Optical parameters such as transmittance can be compared with fluorescence or chlorophyll measurements. If there is an increase in chlorophyll levels for instance, you might expect to see a decrease in transmittance as light is blocked by the layer of chlorophyll. However, if this is not seen in the data the increase in chlorophyll may be a result of instrument error. However, in the same instance it may be the chlorophyll levels which are wrong and it is useful in these situations to check any accompanying documentation for additional information.

Where data series are noisy, perhaps a combined result of turbulent waters and the sensors not being able to adjust fast enough to changing conditions, it is harder to identify erroneous values. In cases like this only the extreme out-of-range data points are flagged, our policy being if unsure it is best not to flag the data point in question (See figure 2).

Comparisons can be made between data from meters on the same rig by overlaying the time series plots on the workstation and by comparing the maximum and minimum values of individual parameters. Similar comparisons are also made between data from neighbouring current meter moorings. Automatic checks are also made to ensure the time channel progresses forwards at equal intervals, problems have been encountered with time channels which jump backwards by a number of hours. In addition, the sampling interval is check in conjunction with the number of data cycles in the series and the start/end time of the series. This is particularly useful when the time channel has not been supplied with the data (Rickards, 1989).

Scatter Plots

Another useful tool provided by SERPLO is the ability to produce scatter plots of the current vectors. This plot is intended for the first appraisal of current meter or other flowfield data. It plots velocity eastings against northings. Typically there are a string of current meters deployed one above the other on a mooring in the water column and the scatter plot allows you to find out more about the following: the extent and orientation of the current ellipse, the presence of outlying points and the relative size of the respective plots. They can be used in conjunction with the time series plot to check that outliers have been flagged.

These plots can also show irregularities in the data, mainly as a result of mechanical malfunction. Examples of malfunctions can be shown up as larger-than-anticipated holes, abnormal symmetry in the tidally-dominated regimes, gaps where a range of speeds or directions are not registered due to meter malfunction, or preferential directions where the compass was not functioning correctly. A typical scatter plot should show symmetry as tides often have regular patterns (e.g. diurnal) with regular speed minima and maxima and directions showing cycles of alternating opposing flow. The orientation and rotation of the tidal ellipse are compared for meters on the same rig and sometimes when possible with meters from neighbouring rigs.

Examples of scatter plots are shown below in figures 4 and 5:

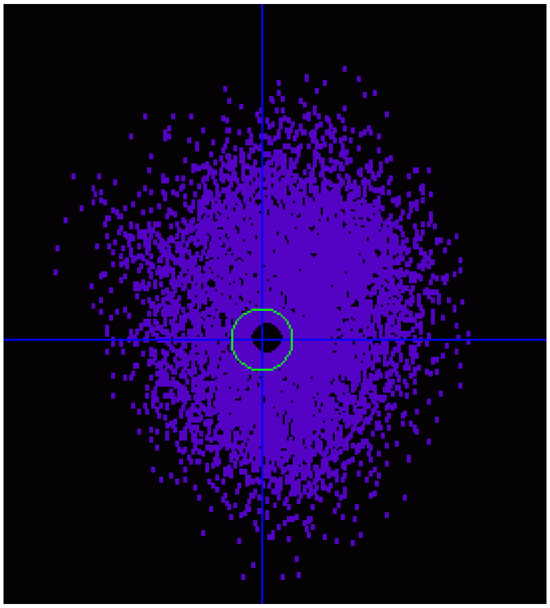

Figure 4 – Example of satisfactory directional data from a data series

Figure 4 shows an example of a ‘good’ scatter plot. There seems to be an even array of current directions indicating that the compass was not being hindered.

However, Figure 5 below is an example of a record with suspect directions as there are very few measurements from approximately 40° to 180°. This specific problem was attributed to the influence of the angle of the Earth’s magnetic field on the compass of an Aanderaa RCM5 current meter.

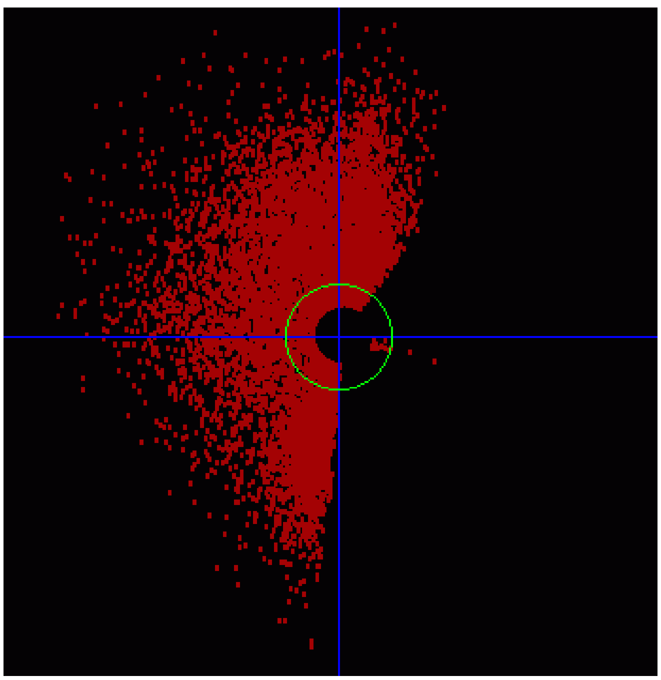

Figure 5 – Example of unsatisfactory directional data from a data series

Common Problems Associated with Current Meters
There are a number of things that can go wrong with a deployed current meter which can affect the quality of the data. This is one of the main reasons why we visually screen data because instrumental errors can be easily picked up. Some of the more common instances when a meter has malfunctioned resulting in a large loss of data are:

- Rotor turns, but there is either a breakdown of magnetic coupling between the rotor and follower or reed switch which then fails to register rotations
- Rotor not turning due to fouling with weed or the suchlike. This results in a sudden drop in speed to zero or near zero.
- Directions not being resolved. This could result from a stiff meter suspension or a meter being fouled by its mooring wire.
- Compass sticking. This may occur if the meter is inclined too far from the horizontal plane and can be a problem in fast tidal streams when in-line instruments are used. This is commonly known as .mooring-knockdown'. This is seen in the data as a frequent recurrence of a single direction value or a narrow range of directions.
- Worn compass. This causes some directions to become repetitive.
- Non linearity of compass. This is usually picked up from the scatter plot of u and v velocity components.
- Sticking encoder pins. This causes spikes in all parameters and is often manifested by the appearance of the value of the pin(s) in the listing (e.g. 0, 256, 512, 768 or 1023).
- Underrated power supply. This often shows in the compass channel first because of the extra current drain during clamping.
- Electronic failure (e.g. dry joints, circuitry broken). This does not always produce a total loss of data however.
- Poor quality recording tape. This is indicated by the appearance of suspect data at regular intervals in all parameters.
- Sensor drift. This is a slow change in the response of the sensor.

Differences in Screening Procedure for ADCPs and Thermistor chains
ADCPs and thermistor chains are different to current meters in that they sample in three dimensions. While ADCPs essentially measure the current in the horizontal plane, they are also able to measure the current at different depths, commonly known as bins, above or below the positioning of the instrument. This is done using the principle of the Doppler shift. An example of an ADCP profile is shown in figure 6 below:

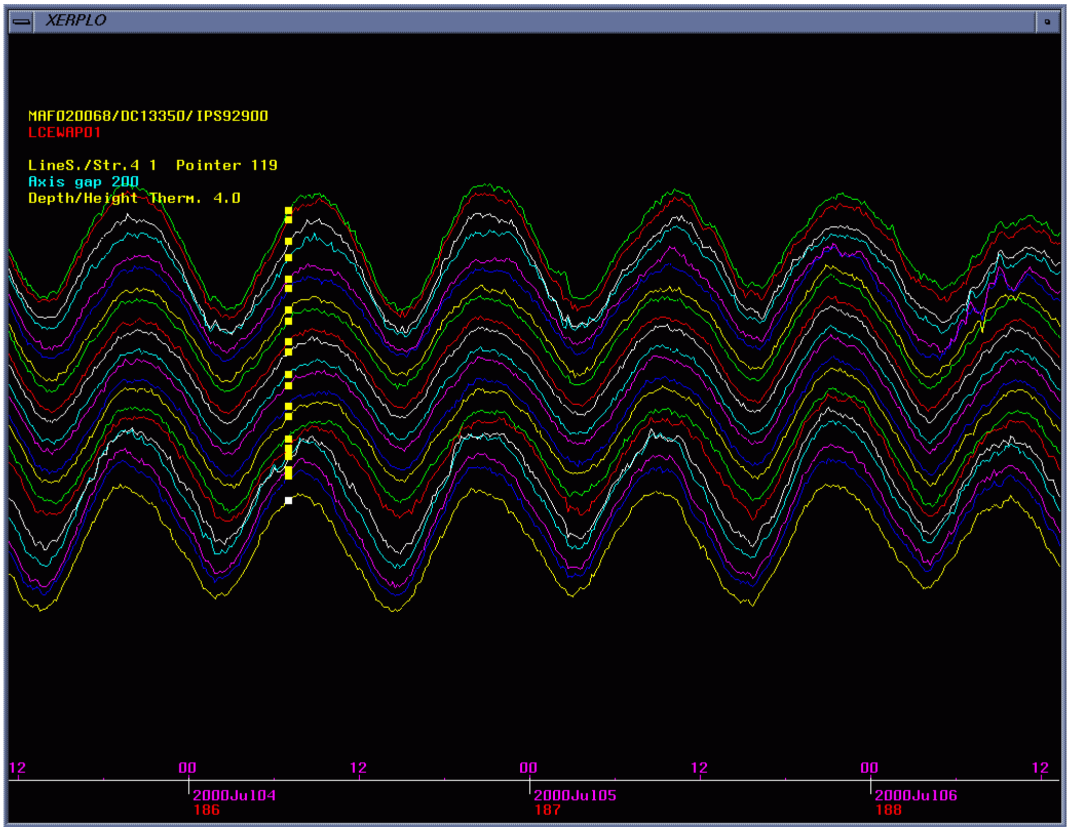

Figure 6 - ADCP profile showing the East-West component of the current direction

Thermistors chains are again similar; however sample temperature at different bin levels.
The screening procedure for these three dimensional data types is similar in principle with that for the normal current meter. The aim is to check for instrument-generated spikes, gaps, repetitive values, and spurious data types at the start and end of the series. Often, the data for one parameter will be observed on the screen at all its measured bin levels. This is useful to show up spikes which may occur up or down the water column. If a spike is present throughout the vertical it may be a valid point due to actual changes in the water properties. However if it is a one-off point in a single bin then the likelihood of it being valid is reduced.

5.3. Wave data

Time Series Plots

Using the time series all parameters can be visually 'screened' with the aim of looking for anomalous values. Often, related parameters, such as significant wave height and maximum wave height, are screened together to identify spurious values. This is because if there is a sudden change in one of the parameters you might expect to see a change in the other in agreement that it is a genuine event.

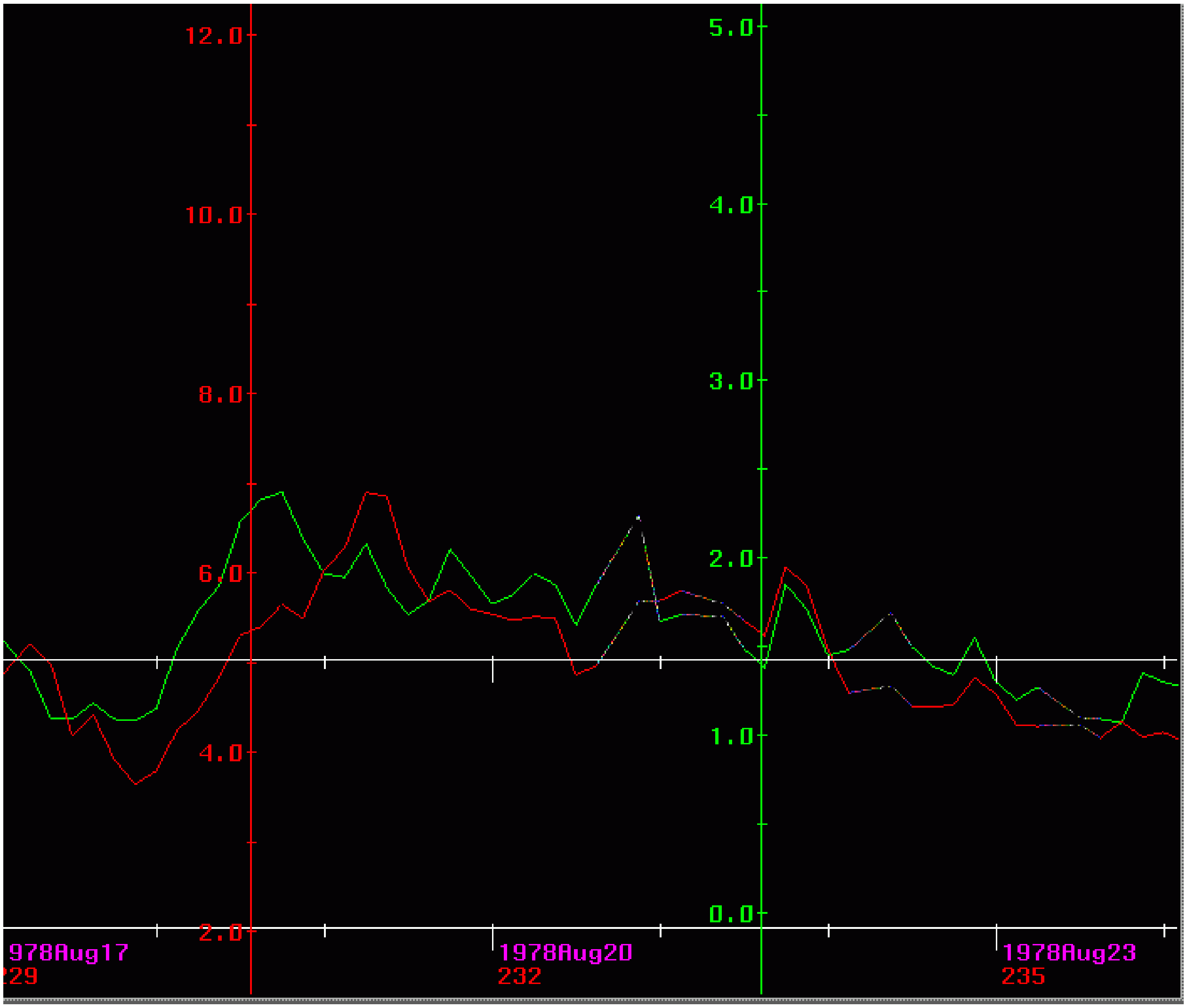

Figure 7 – Time Series of Zero Upcrossing Period and Significant Wave Height

The time series plot shown in this example is that of zero upcrossing period shown in red and significant wave height shown in green.

The time series plots will be used to identify:

- Instrument failure test (10 or more consecutive points of identical value)
- Wandering mean test (interval between successive upcrossings of >25 seconds)

- Check that Tz falls within the range 2-16 seconds
- Check that Tpeak falls within the range 3 -20 seconds
- Check that Tz not less than Tcrest
- Check that Tpeak not less than Tz
- Value in excess of 4 times standard deviation (assumes a basically random process with approximately normal distribution).
- Definition of calm and appropriate flagging
- Changes in wave height/time slopes in excess of 1:10 which is unrealistically steep (NB this is only possible if we have full resolution data i.e. 1Hz or better; unlikely that we will get this, so far data submissions have been 20 minute sampling intervals)
- Check for stationarity: assuming that the wave field is not rapidly evolving or decaying, records of wave height and period should be broadly similar from one record to the next

Scatter Plots
Another useful tool provided by SERPLO is the ability to produce scatter plots of wave height against (zero upcrossing or crest) period, as shown below:

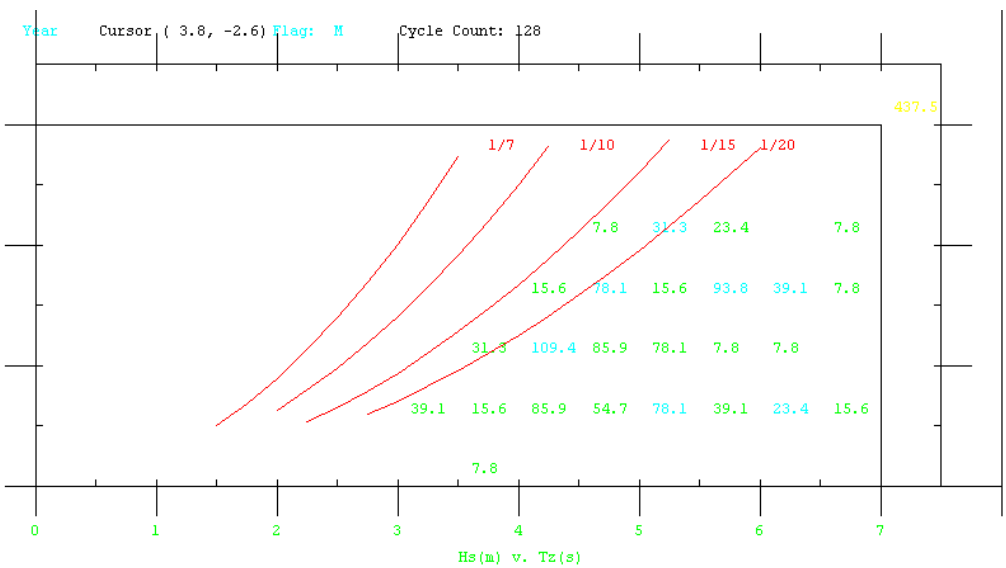

Figure 8 – Scatter plot of wave height against period

These plots can show unrealistically steep waves with a slope of more than 1:10. They can also show outliers from the cluster of Tz vs. Hs values. Similarly, wind speed versus wave height scatter plots can be used to identify outliers from the clusters (N.B. make allowances for swell waves which will show higher than expected wave heights for a given wind speed).

Checks will be made for:

- Definition of unacceptable steepness and appropriate flagging (ratio of Hs/Tz)

- Outliers of clusters of Ts/Hz and of Ts versus wind speed

Frequency Plots
A useful tool provided by SERPLO is the ability to produce frequency distributions of significant wave heights. The tail ends of the distributions can be analysed to identify where instrument noise becomes detectable and a threshold or filter set accordingly.

QC Procedures – 1D and Directional Wave Spectra

- Check slope of energy density spectrum – should follow a set slope due to transfer of energy from lower to higher frequencies (?)
- Check that energy in the spectrum at frequencies below 0.04 Hz is not more than 5% of the total spectral energy
- Check that energy in the spectrum at frequencies above 0.6 Hz is not more than 5% of the total spectral energy
- Check mean direction at high frequencies, which should correspond to the wind direction (assuming coincident meteorological data).
- For 1D spectra, calculate zeroth spectral moment from spectral variance densities and check that it corresponds to the given value
- For 1D spectra, calculate Te as the zeroth divided by first negative spectral moment and check that it correlates with (peak or zero upcrossing) period

Problems Associated with Wave Data
The main problems are expected to be constant values and possibly wandering means, which will be identified as above.

5.4 Sea level

5.4.1 Delayed mode or 'scientific' quality control

As part of the "scientific" or delayed mode quality control (L2), more detailed processing of sea level data is performed, applied to longer time series (typically 1 year) that include not only the steps described for L1, but also filtering to hourly values, computation of annual harmonic constants, residuals, extremes and means. The results of this process are themselves useful products from the station, but also the examination of their quality is crucial for the detection of problems and malfunction in the tide gauge. The primary quality control of sea level is based on the inspection of both recorded data and residuals (i.e. observed – predicted values); inspection of residuals is especially useful for detecting instrumental faults such as timing errors, datum shifts and spikes.

On the other hand, the harmonic constants may be severely corrupted if the site is characterized by highly nonlinear tides, influence of rivers or estuaries and particularly complex basin configuration. To produce more accurate predicted tides, it is advisable to compute 'fresh' tidal constants from recent data and not simply rely upon historical values. Tidal analysis can be performed by means of the software packages developed by the University of Hawaii Sea Level Center (UHSLC) and Puertos del Estado (PE), Spain, that facilitate the use of the Foreman tidal analysis and prediction programs of the Institute of Ocean Sciences, Victoria, British Colombia (Foreman, 1977) or by the PSMSL/Proudman Oceanographic Laboratory (POL TASK2000 Package) and the Australian National Tidal Facility, which utilise the TIRA tidal analysis programs (Murray, 1964). **Tidal constants used in tide predictions should never be mixed between different packages.**

Suspect tidal profiles should be checked against records of a nearby site (also known as "buddy checking"), to see if the suspect values are due to a tide gauge fault or to station conditions. In case of a fault, data should be corrected or interpolated (if possible), otherwise must be maintained unchanged and the event noted. If possible, more than one sensor should be operated at the same site in order to allow direct comparison, and on occasion to fill gaps.

Filtering to hourly values

Raw data are normally registered at time intervals between 1 minute and 1 hour, the most common being 5, 6 and 10 minutes; only in regions where "seiches" occur frequently, or where phenomena such as tsunamis are to be detected, are the sea level registered at less than 1 minute intervals. Apart from the convenience of keeping higher frequency signals for other purposes, it is always necessary to obtain filtered hourly values before proceeding with the sea level processing.

The filtering process will eliminate higher frequencies depending on the frequency cut-off. Pugh (1987) describes useful filters that can be applied to the sea level data at intervals of 5, 10 or 15 minutes to obtain the hourly heights whilst preserving the tidal phenomena. In Godin (1972) there is an extensive discussion on tidal filters.

A test was performed by Puertos del Estado and Instituto Español de Oceanografía using two different filters and data resulting from applying the Godin show a significant lower tidal amplitude. Figures 4 and 5 show that the Godin filter eliminates part of the semidiurnal frequency from the original values while the Pugh filter just eliminates the noise. The Foreman tidal analysis software has taken this into account with a procedure to recover the original amplitude if hourly data have been obtained with the Godin filter. In any case, this is something to be taken into account when comparing hourly sea levels from different institutions or if the computation of extremes or sea level ranges is done from the hourly values (not recommended for this reason).

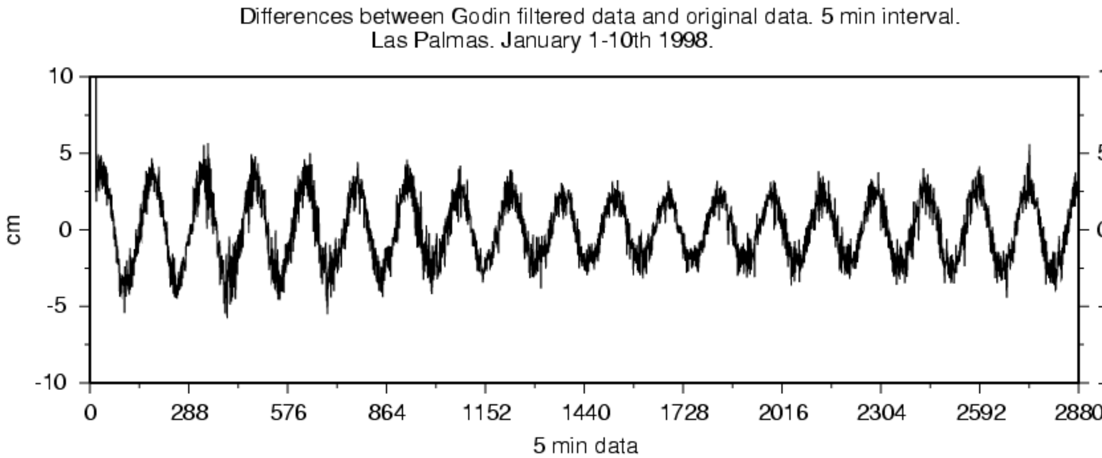

Figure 4: Differences between filtered hourly values and original data after applying the Godin filter show a clear semidiurnal oscillation. Differences in this case oscillate between +- 5 cm.

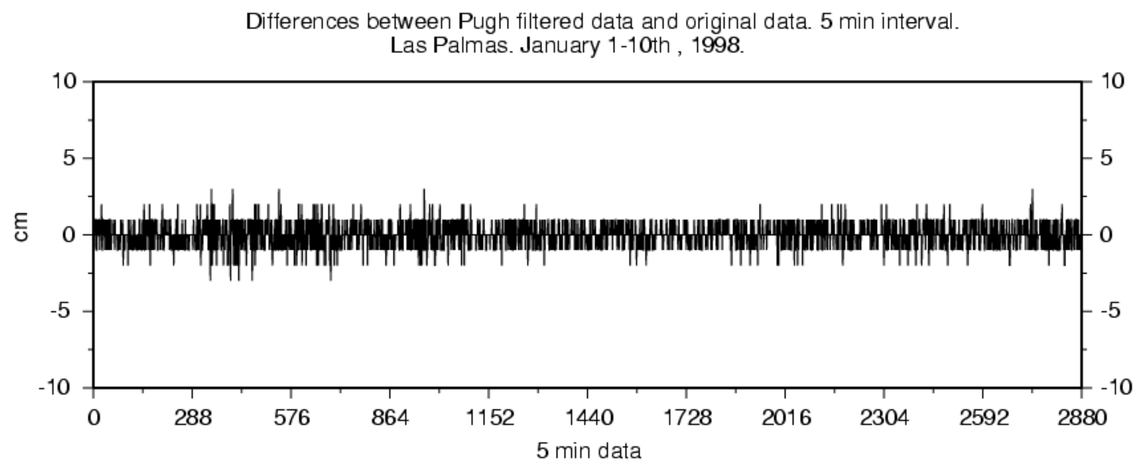

Figure 5: Differences between original data and hourly values for the Pugh filter show clearly that only the noise is eliminated, but no part of tidal signals.

<u>Harmonic analysis</u>

A common procedure is to compute the harmonic constants for each year of observed data. Some of these constants may be particularly affected by meteorological conditions, and so will show important variations from one year to the next. This occurs for example for the longer term harmonic constituents such as Sa and Ssa. Sometimes also the presence of problems in the data series appears as strange values of the normally stable harmonics (e.g. clock errors). In any case, an inspection of the variation through the years of the harmonic constants is interesting both for detecting problems and also for having information about changes on the station. For example, changes in the configuration of a harbour can affect the tide parameters, and this occurs very often.

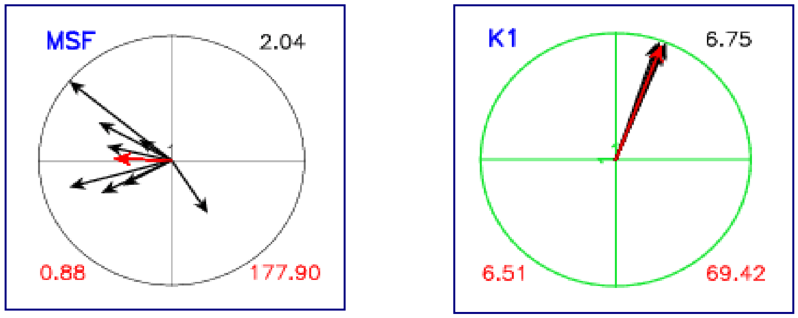

Figure 6: The vector representation is useful to observe the annual variations of the harmonic constants; in black the arrows representing the amplitude and phase of each year for the constituent; in red the mean vector for the vectors of the different years and the values of mean amplitude and phase.

A common practice in order to choose adequate harmonic constants for tide prediction is to perform the vector mean and statistics of the annual values for several years (provided they are computed for nearly complete years and so the same number of constituents have been resolved) and selecting for prediction only the mean of those constituents which do not present a variability that is over a fixed and reasonable tolerance.

A very useful representation is to plot the harmonic constants as vectors, as can be seen in Figure 6 above.

Computation of residuals

As mentioned before, the inspection of residuals is a very useful tool for the quality control process. All fundamental types of errors that a sea level series can present are easily detected in the residual plot. An example of the presence of a clock malfunction (oscillations in the residuals) and a reference change can be observed in Figure 7.

Of course, the presence of a spike is also very obvious in the residual series, which is why some of the automatic algorithms for the detection of spikes are based on both the original and the residual data.

Correction of clock malfunction

This type of error is very easy to correct if there is a constant time shift. The problem becomes evident when there is a drift in the lag between observed and predicted tide.

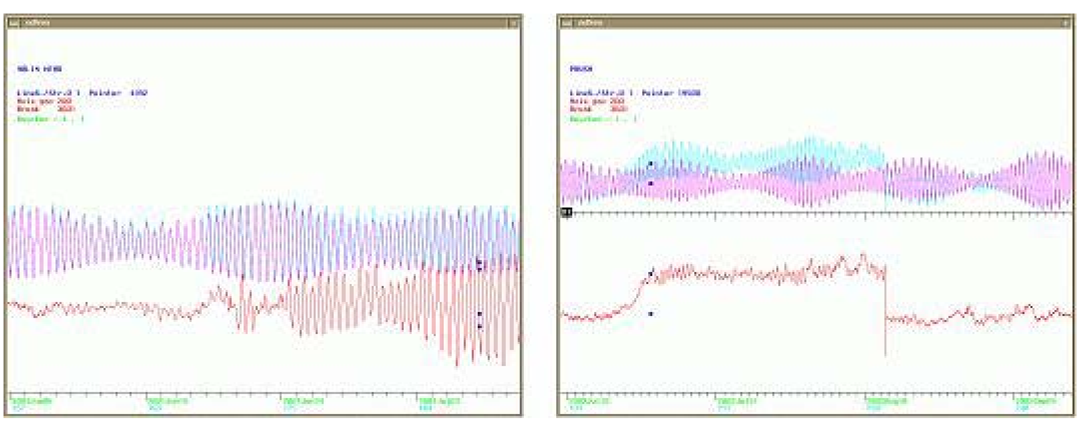

Figure 7: Example of the trace of a clock malfunction and a reference jump in the residuals.

Apart from the inspection of the residuals, a constant lag can be exactly determined by means of lag-correlation analysis between observed and predicted data (lag of maximum correlation), or by comparing the values of the phase of M2 harmonic before and after the shift. Once determined, the part of the series that it is affected must be shifted accordingly to correct the error; if the lag is a multiple of the time interval, the shift is just a movement of data in time; if not, an interpolation to the correct time has to be performed.

This type of correction is not automatic for any of the software currently available, although the UHSLC Sea Level Software includes a program which can make the correction if the lag is a constant value and is a multiple of the time interval.

Gap filling

Depending on the application, filling gaps in a series may or may not be reasonable. During the first stage of quality control of the higher frequency data, very short gaps of several minutes can be linearly interpolated and single point spikes removed and the resulting gap interpolated. Gap filling for hourly values is less clear. The UHSLC interpolates gaps of less than 24 hours before computing daily and monthly means; this is done by computing the residual series, linearly interpolating by using the residual values at the extremes of the gap, and adding on the astronomical prediction to the interpolated values. The maximum length of data that is reasonable to fill should not be more than 24 hours depending on the station. Interpolation of this kind should be undertaken with great caution, and the data values flagged accordingly.

Detection of reference changes

Improper maintenance operation, an accident, or even a natural phenomenon such as an earthquake may produce a sudden jump in the reference level. Most of these jumps are readily identifiable in hourly residual plots if the magnitude is large enough. Once detected, a proper way of correction is through the inspection of the scatter diagram of the tide staff or electric sound readings and corresponding tide gauge values, taken during the maintenance campaigns.

As it will be shown later, these jumps can also be detected by plotting the differences between daily and monthly means from adjacent stations or from redundant sensors. As a rule, the UHSLC considers changes greater than 1.5 cm as significant and worthy of an investigation to guarantee level stability.

A change in reference level must only be corrected and documented when firm confirmation has been established. Data values should be flagged accordingly.

Statistics

Basic statistics from historical data are computed or updated annually and some of these parameters are used for the quality control process. For example:

- upper and lower limits or historical extremes (for range check).
- tidal and observed sea level ranges
- extremes, mean and standard deviation of hourly values, meteorological residuals, ranges or mean sea levels
- tables of monthly and annual extremes
- density function for hourly values, tide predictions and residuals

5.4.2 Further quality control and processing of historical data

When working with historical data, even if the station is well documented, check sheets may not be available with which to perform a confident quality check on the reference level. Furthermore, system measurement problems, changes in the instrumentation or in the environment surrounding the station can generate a discontinuity, which may appear as a datum shift or a trend. In this case some additional checks should be performed to obtain a unique reference. The normal procedure for this kind of higher level quality control is to work with several daily or monthly means sea level series from nearby stations and then reconstruct the time series of the heights.

Different algorithms are explained below that can help to detect this type of discontinuity or reference problems in historical data. All of them normally require the quality assessment of an expert before taking the final decision to correct the data. Apart from the more immediate computation of differences between levels of adjacent stations, which may clarify about the existence of a problem, there are other possibilities as described below.

Correlations

Correlations can be computed both between data from different stations or sensors and between different parameters at the same station (wind, atmospheric pressure, etc). In any case this is a valuable tool for detecting problems. The correlation analysis is also useful for filling gaps. This can be done as follows:

1. Calculate the Pearson's correlation coefficient between residual series
2. Select nearby stations with correlation coefficients above 0.7
3. Calculate the linear regression between them and fill the gaps. (Only fill gaps within the time series; not at the beginning or end of the series)

Standard Normal Homogeneity Test

Several tests have been described in the literature, which can be used to detect inhomogeneities in data series. Alexanderson (1986) developed the Standard Normal Homogeneity Test (SNHT) which is widely used in climatic time series studies. The SNHT gives the points where an inhomogeneity exists and provides information about the probable break magnitude. However, the inhomogeneity could be due to an error or to an anomalous, but real, behaviour of the variable. For this reason, the series are only corrected following comparison with other series in the same climatic region and supported by historical information about the incidences on the tide gauge.

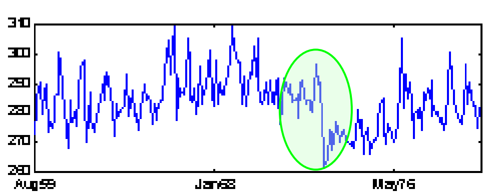	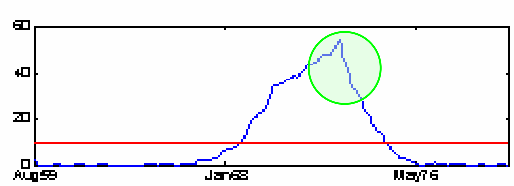
Original series.	SNHT for the original series. Tmax indicates the shift that can be seen in the original series. Red line is the confidence level.

Figure 8: Time series and the result of SNHT

EOF Analysis

The Empirical Orthogonal Functions (EOFs) analysis applies to a group of time series stations can be used not only to find special coherent signals or regional variability but also to detect possible errors in the time series. In fact, relevant differences on the variance of the first EOF may indicate errors in one or more time series. This technique is well documented in “Development of a Quality Checked Tide Gauge Data Set (A.G.P Shaw, M.N. Tsimplis, et all)” and in “Consistency of long sea-level time series in the northern coast of Spain (M. Marcos, D. Gomis, et. al.)

5.5 Chemical sample data (nutrients, oxygen)

A recent IODE workshop on quality control (QC) of chemical oceanographic data collections held at the IOC Project Office for IODE in Oostende, Belgium between 8 and 11 February 2010 met with the objective of evaluating existing procedures and define a minimum set of QC tests and criteria for dissolved inorganic nutrients (phosphate, silicate, nitrate+nitrite, nitrate, nitrite, and ammonium) and dissolved oxygen in seawater (IOC Workshop Report No. 228). The meeting issued a number of recommendations which will be taken forward in post-workshop activities in consultation and in interaction with the wider international community. These included: (1) metadata terminology for reporting measured variables and their units as well as (2) a work plan to recommend a minimum set of numerically defined QC tests that could be adjusted to reflect broad regional to basin scales conditions. These guidelines and recommendations will be assembled on the GE-BICH wiki for peer-review before being published as a technical white paper or guideline document.

Using the GTSPP quality control checks as a starting point, the workshop recommended the following four quantifiable data QC checks for variables as a minimum:

(1) data range checks;

(2) excessive gradient;

(3) excessive spike;

(4) no gradient.

In this context, quantifiable tests are a mean to assign a metric to qualify measured data quality. These tests need to be applicable to both vertical profiles and time-series/horizontal distributions. They are described in detail in the workshop report. The group noted that at present there is no simple way to determine the accuracy of nutrient and oxygen data already available at data centres without independent means (e.g., using certified reference material for nutrients and oxygen data).

In addition to the minimum outlined above, the group recommended the following:

- to develop data ranges at different spatial scales ranging from basin to regions to improve QC checks (objective tests).
- to use property-property plots such as nitrate versus phosphate, nitrate vs. temperature, oxygen vs. temperature, depth or time plots as a mean to visually check the data (subjective tests).
- when available we recommend the use of local climatologies as a quantitative or visual guide.

- to seek community-wise effort to develop less subjective tests.
- to compile a list of available statistical and objectively analyzed climatologies available such as Brickman and Petrie 2003, World Ocean Atlas 2009, Hydrobase, GLODAP, Gouretski and Koltermann 2004, the Baltic Atlas of long term inventory and climatology, Oceanographic Atlas for the Black and Azov seas 2009, MEDAR/MEDATLAS, etc.

5.6 Biological data, etc.,

Biological data brings with it an increased level of complexity, where much attention needs to be paid to the accompanying metadata. There are three papers worthy on note for quality control of biological data:

O'Brien, T.D. 2005. *COPEPOD: A Global Plankton Database.* U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-F/SPO-73, 136 p.

Chapman, A. D. 2005. *Principles of Data Quality*, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

Chapman, A. D. 2005. *Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data*, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

5.6.1 COPEPOD: A Global Plankton Database (2005)

The NMFS Technical Memorandum, *COPEPOD-2005*, summarizes the data processing, quality control, and content of the NMFS-COPEPOD plankton database and notes that:

- Plankton data are variable by nature, influenced by numerous physical and biological events.
- Unlike temperature or salinity values, no tight range of typical values that one can use to easily qualify or disqualify these data.
- Plankton values greatly affected by size of net mesh and depth of tow.
- Very basic value range and statistical techniques to look for anomalous or non-representative data.
- The variety of original units still do not allow for easy inter-comparison of the data: ***Common Baseunit Value (CBV)*** was calculated
- ***Biological Grouping Code (BGC)*** identifies the plankton taxa's membership in up to four groupings

5.6.2 Range checks for biological data (as used for COPEPOD database)

- CBV and BGC are used together to perform broad, taxonomic group-based value range checks
- A single range (for the entire world ocean) was used for the major and minor taxonomic groups.

- Future work will divide these ranges into smaller taxonomic sub-groups and individual oceanographic basins or regions, allowing for tighter range checks
- Value ranges very general and encompass the effects of:
  - Different mesh sizes
  - Day versus night sampling
  - Presence of smaller life stages (“number of adults” vs. “number of adults + juveniles”)
- Will be adjusted as new data and better techniques added to database
- New ranges, as well as ranges for additional plankton sub-groups will be available online.

5.6.3 Statistical checks for biological data (as used in COPEPOD database)

- Used to search for questionable values
- Not used to automatically flag values
- For each BGC group mean and standard deviation calculated based on all observations present in database
- Individual observations >5 standard deviations from mean investigated on a case by-case basis
- Natural variability may account for many “outliers”, method helped identify extreme values caused by misinterpreted units or typographic errors
- In many cases, values off by a factor of 1000
- Readily detected by these simple statistical checks

5.6.4 Chapman papers:

The papers by Chapman were commissioned by the GBIF DIGIT programme to highlight the importance of data quality as it relates to primary species occurrence data. Uunderstanding of these issues and the tools available for facilitating error checking and cleaning is rapidly evolving.

There are many data quality principles that apply when dealing with species data and especially with the spatial aspects of those data. These principles are involved at all stages of the data management process. A loss of data quality at any one of these stages reduces the applicability and uses to which the data can be adequately put. These include:

- Data capture and recording at the time of gathering,
- Data manipulation prior to digitisation (label preparation, copying of data to a ledger, etc.),
- Identification of the collection (specimen, observation) and its recording,
- Digitisation of the data,
- Documentation of the data (capturing and recording the metadata),
- Data storage and archiving,

- Data presentation and dissemination (paper and electronic publications, web-enabled
- databases, etc.),
- Using the data (analysis and manipulation).

All these have an input into the final quality or "fitness for use" of the data and all apply to all aspects of the data – the taxonomic or nomenclatural portion of the data – the "what", the spatial portion – the "where" and other data such as the "who" and the "when".

6. Quality flags

A quality flag is assigned to each data value. Within the SeaDataNet community, the quality flag scheme from the GTSPP project was already in use at a number of data centres. The QC flag scale adopted by SeaDataNet is based on this, with some additions. It has been implemented as one of the SeaDataNet vocabularies (see table below). Quality flags are used to describe the data value; no changes are made to the data values.

Key	Entry Term	Abbreviated term	Term definition
0	no quality control	none	No quality control procedures have been applied to the data value. This is the initial status for all data values entering the working archive.
1	good value	good	Good quality data value that is not part of any identified malfunction and has been verified as consistent with real phenomena during the quality control process.
2	probably good value	probably_good	Data value that is probably consistent with real phenomena but this is unconfirmed or data value forming part of a malfunction that is considered too small to affect the overall quality of the data object of which it is a part.
3	probably bad value	probably_bad	Data value recognised as unusual during quality control that forms part of a feature that is probably inconsistent with real phenomena.
4	bad value	bad	An obviously erroneous data value.
5	changed value	changed	Data value adjusted during quality control. Best practice strongly recommends that the value before the change be preserved in the data or its accompanying metadata.
6	value below detection	BD	The level of the measured phenomenon was too small to be quantified by the technique employed to measure it. The accompanying value is the detection limit for the technique or zero if that value is unknown.
7	value in excess	excess	The level of the measured phenomenon was too large to be quantified by the technique employed to measure it. The accompanying value is the measurement limit for the technique.
8	interpolated value	interpolated	This value has been derived by interpolation from other values in the data object.
9	missing value	missing	The data value is missing. Any accompanying value will be a magic number representing absent data.
A	value phenomenon uncertain	ID_uncertain	There is uncertainty in the description of the measured phenomenon associated with the value such as chemical species or biological entity.

Table ?? SeaDataNet measurand qualifier flags (L201)

7. Documentation

It is important that all data sets are fully documented to ensure they can be used in the future without ambiguity or uncertainty. These should be compiled using, for example, information supplied by the data originator (e.g. data reports, comments on data quality), instrument details, mooring details, data quality, calibration and processing carried out by the data originator and data centre processing and quality control, including any further information gained during QC.

Accompanying Documentation
At BODC we include a set of standard documentation with every data series. Using inhouse software this documentation is 'linked' to the data series in Oracle and when data is requested the accompanying documentation is provided. Example documentation can be found in Annex 2. The most common documents written for each dataset is as follows:

Data Activity – this document describes the 'event'. In the case of wave buoys or similar it generally gives the dates of deployment, description of the location including the latitude and longitude. Any other information to the event is included in this document. This document is linked to every data series which has been collected from the same mooring.

Data Quality Document – this document is linked to individual series and is not generic. Any comments or problems relating to the data series are included in this document as well as any steps taken to resolve the problem. Often this is provided by the data originator where they have taken steps to improve the quality of the data and our input is generally made up of comments from the screening process. Documentation can come under the following headings:

Project Report – This is a document describing the project for which data is collected. This is a generic document which is linked to every data series involved with a particular project.

Fixed Station Document – This gives information on a particular station which is used consistently for measurements.

Restrictions Document – This document outlines any restrictions imposed on particular datasets and who the main contact is for any questions relating to the data.

Screening Document – This is a generic document linked to all datasets giving a brief summary of the screening procedure undertaken for each dataset, so the external user is aware of the broader quality control that takes place.

Instrument Document – This is a generic document linked to every series which has been produced from a particular instrument. The document includes information on how the instrument works, its sensitivity, accuracy and links to the manufacturer's website where applicable.

References

NODC procedures (e.g. France, Greece, Italy, Norway, Spain, Sweden, UK)

EU MEDAR-MEDATLAS procedures and SCOOP software

EU SIMORC project (Met-ocean data QC)

EU ESEAS (sea level) and IOC GLOSS documents

Manual of Quality Control Procedures for Validation of Oceanographic Data, UNESCO, IOC - Manuals & Guides, 1993, Manual And Guides 26

GTSPP Real Time Quality Control (IOC Manuals and Guides No. 22) Second Revised Edition

Argo Quality Control Manual (Real Time and Delayed Mode)

GOSUD Real-time quality control

IODE's OceanTeacher

ICES WG Data and Information Management Data Type Guidelines

JPOTS Manual, 1991

WOCE manuals

JGOFS Protocols

World Ocean Database Quality Control documentation

TOGA/COARE Handbook of Quality Control Procedures for Surface Meteorology Data

BODC-WOCE Sea Level Data Assembly Centre Quality Assessment

AODC Quality Control Cookbook for XBT Data

MyOcean Real Time Quality Control of temperature and salinity measurements

INS TAC: Recommended Quality Control on Sea level *in-situ* data within MyOcean

MyOcean Real Time Quality Control of Current measurements

MyOcean Quality control procedure on current and temperature measurements derived from drifters

MyOcean Real Time Quality Control of biogeochemical measurements

Chapman, A. D. 2005. *Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data*, version 1.0.

Chapman, A. D. 2005. *Principles of Data Quality*, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

‘Ocean biodiversity informatics’: a new era in marine biology research and management (Mark J. Costello, Edward Vanden Berghe)

O’Brien, T.D. 2005. *COPEPOD: A Global Plankton Database.* U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-F/SPO-73, 136 p.

QARTOD (Quality Assurance of Real-Time Oceanographic Data)

UNESCO. 1988. Guide to Drifting Data Buoys, IOC/WMO Manuals and Guides # 20.

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

WMO. 1995. WMO Manual on Codes No. 306.

UNESCO. 1991. Manual on International Oceanographic Data Exchange, IOC/ICSU Manual and Guides # 9, Revised Edition.

Emery, W.J. and R. E. Thomson. 1998. Data Analysis Methods in Physical Oceanography, Pergamon.

Karl, David, Luis Tupas, Fernando Santiago-Mandujanu, Craig Nosse, Dale Hebel, Eric Firing and Roger Lukas. 1996. Hawaii Ocean Time-Series Data Report 7:1995, SOEST 96-09, University of Hawaii.

UNESCO. 1988. The acquisition, calibration and analysis of CTD data. A report of SCOR WG 51. Tech. Pap. Mar Sci., 54: 59pp.

UNESCO, 1991. Processing of Oceanographic Station Data, JPOTS Editorial Panel.

First IODE Workshop on Quality Control of Chemical Oceanographic Data Collections, IOC Project Office for IODE, Oostende, Belgium, 8-11 February 2010 Paris, UNESCO, 25 March 2010 (IOC Workshop Report No. 228) (English)

Appendix 1 ICES Data Type Guidelines

ICES Guidelines for XBT Data

(Compiled January 2000; Revised August 2001; April 2006)

The Expendable Bathythermograph (XBT) has been used by oceanographers for many years to obtain information on the temperature structure of the ocean to depths of up to 1500 meters. The XBT probe is typically launched from a steaming ship. During the probes descent, it measures the water temperature. Two very small wires transmit the temperature data to a ship computer where it is recorded for later analysis. The probe is designed to fall at a constant rate, so that the depth of the probe can be inferred from the time since it was launched.

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.

1.1 Data Standard

An overview of the instrument, data logging practices, data transmission in real-time (low resolution), and the overall data management practices for high resolution data in a continuously managed database is provided in Annex A.

All parameters must be clearly specified and described. If parameter codes are to be used, then the source data dictionary must be specified. 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 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 times of the removed values should be noted.

Sufficient self-explanatory information and documentation should accompany the 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 data supplier should ensure that the following be provided with the XBT data submission:

- Metadata information about the instruments (XBT manufacturer, probe model, probe recording rate)
- Metadata information about the data precision and final accuracy
- The fall rate equation used
- Metadata information about the calibration and processing techniques used for the particular XBT data set
- 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
- Any ancillary meteorological information should be included

1.2 Format Description

Low-resolution XBT data, transmitted in real-time over the Global Telecommunications System free of charge, are coded into the WMO BATHY report format JJVV (the previous format was JJYY. Changes to JJYY that produced JJVV may be found at http://www.meds-sdmm.dfo-mpo.gc.ca/meds/prog%5Fint/j%2Dcomm/codes/new%5Fbathy%5Fe.htm) (see also Annex B). Coding of the messages can be done either manually or automatically via automated data collection platforms, such as the US Shipboard Environmental Acquisition System (SEAS at http://seas.amverseas.noaa.gov/seas), via GOES or INMARSAT satellites or using the French Service ARGOS.

The coded message is typically made up of inflection points from the profile. Various algorithms exist for determining these inflection points (e.g. Rual, 1989).

The originator data formats for the exchange of high resolution XBT data may vary. However, data should be supplied in a fully documented ASCII format. Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is strongly recommended. Ideally all of the data from the instrument should be stored in a single file. The contents of the data and ancillary information should adhere to the [Formatting Guidelines for Oceanographic Data Exchange (http://www.ices.dk/ocean/formats/getade_guide.htm)](http://www.ices.dk/ocean/formats/getade_guide.htm) prepared by the IOC's Group of Experts on the Technical Aspects of Data Exchange (GETADE) and available from RNODC Formats.

The high resolution XBT data set must include:

- A full description of the format used.
- All XBT profiles collected from a single platform can be in the same file, or one file can be used for each XBT profile.
- The files must be homogeneous (i.e. each piece of information must always be in the same place in the file).
- Data should be reported at 1 m resolution

1.3 Collection Details

Other pertinent information to be included in the data transfer to the Data Centre includes:

- Platform name
- Station number
- Country, organisation
- Date and time of the start of each profile
- Sounding (specify method, e.g. chart, constant speed of sound, etc.)
- Latitude and longitude of each station, method of position fix (e.g. GPS, DGPS)
- calibration reference temperature and the method used to determine the calibration

Any additional information of use to secondary users which may have affected the data or have a bearing on its subsequent use. An example deployment log sheet is provided in Annex C.

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 XBT 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)](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 high-resolution 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 XBT data by other data centres. The UKHO (1999) guide and Bailey et al. (1994) are also excellent references related to processing raw XBT data from a variety of source instruments.

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

The first component examines the characteristics of the platform 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 data of lower resolution (such as a real-time BATHY message) typically 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.

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 Principal Investigator (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
- Any ancillary parameters (e.g. meteorological data)
- Parameter and unit definitions, and scales of reference
- Definition of flagging scheme, if flags are used
- Relevant information included in the data file (e.g. ship, cruise, project, start and end dates, 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, see below)
- Any problems encountered with the data and their resolution
- Any changes made to the data and the date of the change

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.

3.3 Referral Service

ICES member research and operational data centres produce a variety of data analysis products and referral service. 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

Bailey, R., A. Gronell, H. Phillips, G. Meyers and E. Tanner, 1994: CSIRO Cookbook for Quality Control of Expendable Bathythermograph (XBT) Data. CSIRO Marine Laboratories Report, 221, 75 pp.

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.

UKHO, 1999. XBT Data Processing User Guide, UK Hydrographic Office, June 1999.

Annex A

An XBT probe consists of a small projectile with a leaded nose and a plastic casing that contains a spool of fine copper wire. As the probe descends through the water column, the wire self-spools and transmits the temperature from a thermistor in the nose to a recorder on deck. Corresponding depths are computed from an empirically derived, second order fall rate equation given by the manufacturer of the probe. Historically, a continuous analog chart was used to record XBT temperature-depth readings, which were subsequently digitised. More recently digital recorders have replaced these analog charts.

The probes are generally launched by research ships, naval vessels, and by the crew aboard volunteer merchant ships that transit regular shipping routes. Sippican Inc. (http://www.sippican.com/expendable_probes.html) has prepared an XBT User's Guide that gives practical advice and hints on how to successfully store and deploy XBTs. The guide is based on deployment experiences with XBTs over many years.

The term high-resolution is used for profiles that are sampled usually at 1 or 2 metre depth resolution, require maximum calibration and quality control, and hence are normally available in delayed mode. Low-resolution messages are those profiles that have been reduced by some process such as selecting inflection points to reproduce the trace to specific tolerances (Rual, 1989). Some data centres maintain archives of only the reduced profiles. As well, the reduction tolerances may vary from agency to agency. Some of these messages are transmitted and available in real-time (less than one month, usually within days of observation). Delayed mode high resolution data are more useful to scientists but generally it is a matter of months to years before the high-resolution data are submitted to data centres.

Requirements for the collection and exchange of low-resolution XBT messages for real-time operations differ from those of an oceanographic data archive centre for high-resolution, fully calibrated and quality controlled XBT data. For example, although XBTs have been used for some time, it is only in the last few years that errors have been detected and corrected in the fall-rate equation coefficients and with the thermistors. Problems have also been detected in shipboard XBT recorders. For these reasons, there is a higher requirement for metadata information such as instrument type and the associated calibration information along with the original recordings.

To manage XBT data, it is recommended that 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 messages 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.

Annex B

HOW TO PREPARE A JJVV BATHY REPORT WITH NEW COEFFICIENTS

As of May 3, 2000 the WMO bathy report format JJYY was replaced with the format JJVV shown below.

Code Table A lists codes for IxIxIx and Code Table B lists codes for XrXr.

Example: On November 29, 1995, at 13:45 GMT, a vessel underway deploys an XBT using a Sippican T-7 probe with coefficients a=6.691 and b=-2.25 (the new coefficients have been programmed into the shipboard software) and a Sippican MK-9 recorder. The probe was deployed at 45 25'N and 150 37'W. The depth-temperature inflection points are selected automatically by the recorder.

BATHY Code:

Section 1

MiMiMjMj	YYMMJ	Gggg/	QcLaLaLaLaLa	LoLoLoLoLoLo
JJVV	29115	1345/	74525	15037

Section 2

8888k1	IxIxIxXrXr	ZoZoToToTo	.....	ZnZnTnTnTn
88887	04203			

CODE INPUT:

1. MiMiMjMj represents the report and version, JJVV, denoting a BATHY report encoded after 3 May 2000.

2. Date and time are given in GMT (YYMMJ and Gggg/).

3. Qc = 7 = quadrant of the glove; from Table 3333 of WMO No. 306.

4. Latitude and longitude given in tenths, hundredths and thousandths of a degree, depending on the capability of the positioning system. When the position is in tenths of a degree, the position group shall be encoded as $Q_cL_aL_aL_a$// or $L_oL_oL_oL_o$//. When the position is in hundredths of a degree, the latitude position group shall be encoded as $Q_cL_aL_aL_aL_a$/ or $L_oL_oL_oL_oL_o$/; iudff and 4SnTTT are optional fields for wind direction and speed and air temperature.

5. 8888 indicates data on Instrumentation and Temperatures at either significant or selected depths follow.

6. k1 = 7; from table 2262 of WMO No. 306 (auto-selection of data points).

7. IxIxIx = 042 (Sippican T-7,6.691,-2.25);from Code Table A below.

8. XrXr = 03 (Sippican MK-9 recorder); from Code Table B below.

9. ZoZo.......ZnZn = significant depths, in metres, starting with the surface.

10. ToTo.......TnTnTn = temperatures, to tenths of a degree Celsius, starting at the surface.

CODE TABLE A for IxIxIx - Instrument Type for XBT with fall rate equation coefficients.

Code Instrument, equation coefficients a, and b

001	Sippican T-4, 6.472,-2.16
002	Sippican T-4, 6.691,-2.25
011	Sippican T-5, 6.828,-1.82
021	Sippican Fast Deep, 6.346,-1.82
031	Sippican T-6, 6.472,-2.16
032	Sippican T-6, 6.691,-2.25
041	Sippican T-7, 6.472,-2.16
042	Sippican T-7, 6.691,-2.25
051	Sippican Deep Blue, 6.472,-2.16
052	Sippican Deep Blue, 6.691,-2.25
061	Sippican T-10, 6.3301,-2.16
071	Sippican T-11, 1.779,-0.255
201	TSK T-4, 6.472,-2.16
202	TSK T-4, 6.691,-2.25
211	TSK T-6, 6.472,-2.16
212	TSK T-6, 6.691,-2.25
221	TSK T-7, 6.472,-2.16
222	TSK T-7, 6.691,-2.25
401	Spartan XBT-1, 6.301,-2.16
411	Spartan XBT-3, 5.861,-0.0904
421	Spartan XBT-4, 6.472,-2.16
431	Spartan XBT-5, 6.828,-1.82
441	Spartan XBT-5DB, 6.828,-1.82
451	Spartan XBT-6,6.472,-2.16
461	Spartan XBT-7,6.472,-2.16
471	Spartan XBT-7DB, 6.472,-2.16
481	Spartan XBT-10, 6.301,-2.16
491	Spartan XBT-20, 6.472,-2.16
501	Spartan XBT-20DB,6.472,-2.16
800	Mechanical BT
810	Hydrocast
820	Thermistor Chain
830	CTD

Notes:

(1) The depth is claculated from coefficients a and b and the time, t, as follows: z = at + 10-3bt2.

(2) All unassigned numbers are reserved.

CODE TABLE B for XrXr - Recorder Types

This table encodes the various recorders used to log temperatures from the instruments listed.

Code	Recorder
01	Sippican StripChart Recorder
02	Sippican MK2A/SSQ-61
03	Sippican MK-9
04	Sippican AN/BHQ-7/MK8
05	Spiican MK-12
10	Spartan SOC BT/SV Processor Model 100
20	Argos XBT-ST
21	CLS-AGROS/Protecno XBT-ST model 1
22	CLS-ARGOS/Protecno XBT-ST model 2

30	BATHY Systems SA-810
31	Scripps Metrobyte controller
32	Murayama Denki Z-60-16 IIIjMj
33	Murayama Denki Z-60-16 II
34	Protecno ETSM2
35	Nautilus Marine Service NMS-XBT
40	TSK MK-2A
41	TSK MK-2S
42	TSK MK-30
43	TSK MK-30N
99	Unknown

Annex C
Example XBT Deployment Log Sheet

XBT Deployment Log Sheet

Country:____________ Organisation:______________ Platform:____________
Date:_______________

Event Number	Latitude dd mm.mm	Longitude dd mm.mm	Time (GMT)	Sounding	Filename	Initials

ICES Guidelines for CTD Data

(Compiled March 2000, revised August 2001 ; June 2006)

CTD (conductivity, temperature and depth) instruments were introduced to the oceanography community in the late 1960's. Since then, the electronic measurement of conductivity, temperature and pressure provided by the CTD has become the backbone of hydrography measurements in the ocean.

CTDs typically consist of an array of sensors that measure the frequency or voltage response that represents changes in an ocean parameter. Beyond the typical conductivity and temperature sensors, CTDs may also have attached sensors for light transmission, fluorescence, oxygen content, optical backscatter and turbidity. For details regarding the use of CTDs, see the attached training annex.

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.

1.1 Data Standard

All stations from a cruise can be in the same file (recommended), or one file can be used for each station.

It is recommended that the header information, the configuration file and the raw data file be included in the same file. If this is not possible, then it is acceptable to include the header information in a separate (master) file. In this case, an unambiguous index connecting the header with the data file must be used. This index should preferably be a construction of a station/cast number and should be part of the data file name.

The files must be homogeneous. For the headers, that means that each piece of information must always be in the same place in the file or each data element should be terminated with a field separator. For the data, this means that all of the files (for one cruise) must have the same parameters in the same order. If the data file does not include a definition statement of what data are following in what order then certain rules must be followed. In particular even if one parameter is not measured at one particular station, it can be replaced by its null value, in order to have a fixed number of columns for files of the same cruise. The null value, which must not be confused with valid data, may be a large negative number (e.g. -99.999).

Only the down casts of the station should be provided except if only the upcast is available. If water samples have been collected, discrete upcast CTD values (discrete refers to CTD values collected at the time of bottle trip) should be provided as well.

The file needs to contain the measured parameters, in situ temperature (not potential temperature), pressure (not depth), salinity, etc.

The recommended pressure interval is 1 decibar

All parameters must be clearly specified and described. If parameter codes are to be used, then the source data dictionary must be specified. 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 data including laboratory and field calibrations. The data should be fully checked for quality and flagged for erroneous values such as spikes, gaps, etc. Laboratory suspect flags – different organizations have different approaches to suspect flags for example CEFAS (UK) denotes suspect data with a ‘1’. This system isn’t necessarily used by other organisations or ICES. An explicit statement should be made of the checks and edits applied to the data.

If a cruise/data report is available describing the data collection and processing, this can be referenced. If possible a copy of the calibration record and Roscop form should be supplied with the data.

1.2 Format Description
Data should be supplied in a fully documented ASCII format although the preferred method of transport is ftp. Any large submissions should be compressed using a winzip compatible compression routine. Data Centres are capable of handling CTD data in a wide variety of user-defined and project formats (e.g., WOCE). If in doubt about the suitability of any particular format, advice from the Data Centre should be sought.

Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is strongly recommended. Ideally all of the data from the instrument should be stored in a single file. The contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange (http://www.ices.dk/ocean/formats/getade_guide.htm) 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
The collection of quality CTD data requires considerable training and care. For this reason, we provide here a training annex for those new to CTD operations.

Pertinent information to be included in the data transfer to the Data Centre includes:

- Project, platform, cruise identifier
- Country, organization
- Station number, Site,
- Date and time of the start and end of the sampling
- Position (latitude and longitude degrees and minutes or decimal degrees can be used. Explicitly state which format is being used. It is recommended that N, S, E and W labels are used instead of plus and minus signs.)
- Description of operational procedures including sampling method, sampling rate, sensor resolutions, methods of position fixing (e.g. GPS, DGPS)
- Details of the instrument and sensors (e.g. manufacturer, model number, serial number, and sampling rate)
- Station depth and sample depth should be included for each station. The method and assumptions of determining the sounding should be included.

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
A range of checks are carried out on the data to ensure that they have been imported into the Data Centre s format correctly and without any loss of information. For CTD data, these should include:

- Check header details (vessel, cruise number, station numbers, date/time, latitude/longitude (start and end), instrument number and type, station depth, cast (up or down)), data type/no. of data points)
- Plot station positions to check not on land
- Check ship speed between stations to look for incorrect position or date/time
- Automatic range checking of each parameter
- Check units of parameters supplied
- Check pressure increasing
- Check no data points below bottom depth
- Check depths against echo sounder
- Plot profiles (individually, in groups, etc)
- Check for spikes
- Check for vertical stability/inversions
- Plot temperature vs. salinity

- Check profiles vs. climatology for the region
- Check calibration information available

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 Principal Investigator (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. This is described in more detail below:

- 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 data file (e.g. ship, cruise, project, start and end dates, 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 Any changes made to the data and the date of the change

- 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.

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.

ICES CTD Training Annex

(Compiled March 2000, revised April 2000, ; June 2006)

CTD instruments are capable of measuring conductivity, temperature and pressure to a high accuracy provided they are used correctly. This requires all staff collecting CTD data to receive basic training in how the instrument works and what its capabilities are, how to calibrate the instrument, how to deploy the instrument at sea and successfully log the measurements using instrumentation software, and if necessary, how to subsequently process the profile. In order to achieve accurate measurements considerable care has to be taken, especially in the derivation and application of sensor calibrations, and users should be made aware of potential problems. Guidelines may be useful, but an **understanding of the instrument and procedures to be adopted** is better than following instructions while not appreciating the reasoning behind them. Proper training must be emphasised.

The following outlines the key points to collecting quality CTD data.

A.1 USE CAREFULLY DRAFTED LOGSHEETS

Logsheets are an important requirement when CTD data are being collected and the need for correctly completed logsheets should be emphasised when training those new to collecting CTD data. The logsheets should be drafted while at sea so that any problems or queries arising from any stations can be dealt with promptly.
These sheets go further than recording the station position, time, etc. and if carefully drafted they will prompt users to record much of the relevant information that is needed. This includes for example, the serial number of the instrument, the identity of reversing thermometers used to compare with the CTD temperature sensor, and the identity of the persons who worked the station, read the thermometers and collected the water samples for example, salinity and nutrient analysis. Users should be encouraged to record as much relevant information as possible on the logsheet (further comments box), especially any unusual features such as indications of sensor malfunctions and large wire angles.

A.2 PRE-DEPLOYMENT CHECKS

It is wise to complete an examination of the CTD sensors prior to deployment to check that:

- No fouling of sensors and any protective coverings have been removed
- Lenses of optical sensors are clean
- Thermometers are correctly set (if fitted)
- Bottles are correctly set and taps closed (if used)
- Any additional battery supplies are switched on
- With the instrument switched on observe the values displayed by the sensors to confirm that they are functioning correctly and record the value displayed by the pressure and temperature sensors whilst the CTD is on deck. This can be used to correct the logged pressure.

To avoid confusion, it is helpful if each team adopts an agreed procedure, so that each member will complete the same checks at each station. An itemised check list is useful.

A.3 SENSOR CALIBRATION

To obtain the highest quality data, corrections need to be applied to the CTD sensors.
Calibration procedures will vary from one laboratory to another, but it is generally accepted that whilst the pressure and temperature sensors can be subject to pre- and post-cruise calibrations in the laboratory, the conductivity sensor is best calibrated by comparison with samples collected for salinity analysis. A pressure correction for each station can be determined by noting the pressure when the instrument is on deck, but some pressure sensors are temperature sensitive and a further correction may be necessary. A dead-weight tester is often used to obtain a pressure calibration in the laboratory and the results from this should be in good agreement with the observed 'on deck' value.

In the laboratory, a temperature sensor is readily calibrated by comparing its readings with temperatures from a platinum resistance thermometer and this gives a more accurate calibration than can be achieved with reversing thermometers. However, a thermometer will provide a check on the CTD temperature and may indicate the presence of a temperature 'jump'. Thermometers are also useful to identify rosette misfires.

Care is needed when taking samples for salinity analysis to compare with the CTD conductivity. A standard operating procedure document for this process is recommended. For the highest quality salinity (or conductivity) data corrections, it is recommended that water sample salinity replicates be drawn. Replicates not satisfying some pre-set criteria (e.g. difference of 0.002) may be rejected. It is also very important that the salinometer being used for salinity samples be maintained to a high standard.

Care is also required when using reversing thermometers for comparison with the CTD temperature estimates. Those depths where temperature (or salinity) gradients are known to exist should be avoided. If samples have to be collected from such depths the logsheet should be clearly marked to this effect and it is probably advisable not to include them in the calibration computation. If it is required to sample in the thermocline (or halocline) then it would be wiser to add additional sampling depths suitable for calibration. The operator has to be vigilant when the samples are being collected. If thermometers are being used sufficient time must be allowed for equilibrium (at least five minutes for mercury-in-glass thermometers). See also Karl (1996), UNESCO (1988) and UNESCO (1991).

Throughout this time the operator should be viewing the values displayed by the CTD and if they are variable this should be clearly recorded on the logsheet and preferably not used to determine the calibration coefficients. Many CTD users have the instrument mounted in a multisampler rosette that accommodates bottles, perhaps fitted with reversing thermometers. These bottles are closed in pairs at selected depths to collect the sample for salinity analysis that is eventually compared with the derived CTD estimate. Care has to be exercised when using this data to identify rosette misfires (i.e. when a bottle does not fire at the selected depth). Sometimes a bottle does not respond to the triggering signal or two (or more) bottles close simultaneously. Often, if a misfire takes place all subsequent samples collected during the cast will not be from the intended depth. The actual sampling depths need to be established when deriving the calibration coefficients. That a misfire has occurred is not always obvious when the CTD is returned to the surface, but all users should be made aware that this can (and does) happen and that they must look closely at the data to check for this.

A.4 SENSOR RESPONSE

An important feature of the CTD instrument that causes problems is the mismatch in temperature and conductivity signals due to the time response and physical offset between the sensors. If improperly accounted for in the processing, these differences result in salinity 'spikes'. All users should be aware of this problem and if necessary a procedure for removing them needs to be adopted. The report of the SCOR WG 51 discusses this in detail.

Software packages are now available, often purchased with the instrument, which purport to remedy this problem. It is recommended that a careful appraisal of such packages is made before deciding whether to use them. Local conditions can also influence data quality (e.g. the response of the conductivity sensor in waters with a large sediment load may be impaired).

A.5 POST PROCESSING

It is recommended that some processing of the data be completed at sea, preferably soon after the CTD station is complete. A database can be used for this purpose where all general details, niskin information, digital thermometer, salinity bottles numbers are recorded for each station. This is often the only way of detecting an instrument malfunction (e.g. a noisy sensor) and a comparison between CTD and thermometer temperatures, CTD and salinometer salinity estimates should be made regularly during the cruise. The data from other sensors being logged should also be examined. This 'first look' offers an opportunity to identify samples unsuitable for use in derivation of the sensor calibration coefficients (e.g. varying estimates). It is useful to have the pre-cruise calibration data at sea so that checks on performance can be compared with the most recent laboratory calibrations during the cruise.

REFERENCES

Emery, W.J. and R. E. Thomson. 1998. Data Analysis Methods in Physical Oceanography, Pergamon.

Karl, David, Luis Tupas, Fernando Santiago-Mandujanu, Craig Nosse, Dale Hebel, Eric Firing and Roger Lukas. 1996. Hawaii Ocean Time-Series Data Report 7:1995, SOEST 96-09,
University of Hawaii.

UNESCO. 1988. The acquisition, calibration and analysis of CTD data. A report of SCOR WG 51. Tech. Pap. Mar Sci., 54: 59pp.

UNESCO, 1991. Processing of Oceanographic Station Data, JPOTS Editorial Panel.

ICES Guidelines for Multibeam Echosounder Data

(Compiled September 2006)

If you are considering producing a hydrographic survey please read the following information first. Your survey could be used to compile new or updated nautical charts and so help improve safety for mariners. Your survey does not have to be carried out to full charting standards and all data collected is potentially of use. Even without a full search for dangers, a modern survey can improve on our knowledge of the seabed.

Some advice on how to help us to get the best out of your survey follows below. But even if you are unsure about these requirements, we would still like to receive a copy of your survey. It may reveal changes that should be shown on the chart as a "less water reported" legend, and/or may lead to a full survey being carried out.

Hydrographic and bathymetric data centres such as the UK Hydrographic Office's (UKHO) Seabed Data Centre (SDC) (see Annex B for contact details) are always happy to give help and advice on hydrographic data gathering requirements in order to increase the usefulness of hydrographic survey data.

Just a few minor changes to how your data is gathered and presented can make it much more useful for charting and other purposes.

1.0 Receiving Data

Hydrographic and bathymetric data centres require the following information to be supplied by the data supplier together with the data.

1.1 Data Standard

This guideline is designed for use by agencies working outside of the immediate area of hydrographic surveys for charting. Many agencies collect swathe (i.e. multibeam or interferometric) echosounder data in the course of their operations at sea, and much of this data can often be of great use for navigational purposes, especially in areas of poor survey coverage. Although all such data is of use, and hydrographic offices such as the UKHO would welcome the receipt of it, its usefulness can be greatly enhanced if data is collected to a suitable standard and sufficient information is provided to permit the receiving hydrographic office to assess the data. With this in mind, the following information is provided to help agencies provide data that may be fully utilised in future navigational products.

1.2 Format Description

Capability may vary from office to office. For example the UKHO is able to accept processed swathe data in the following formats:

- Caris HIPS Project
- Kongsberg Maritime "Neptune" Project
- IVS "Fledermaus" PFM
- GSF
- ASCII

Raw Data files

All raw data should be supplied where feasible. It is useful to keep this separate to the processed data. The UKHO is able to accept raw swathe data in a variety of formats, including;

- Kongsberg Maritime "Merlin" format (.all files)
- XTF
- GSF
- Atlas
- Elac
- Geoacoustics
- Hypack
- LADS
- SHOALS

Caris HIPS VCF
If Caris HIPS has been used to process the data, then any VCF or HVF files would be very useful.

Delivery Media
Capabilities will vary but as an example the UKHO is able to accept data on various electronic media (including FTP). The media should be appropriate to the size of the data set.

- For datasets of 20GB or more, we recommend USB2 portable hard disk(s). These can be returned to the sender within 2 weeks.
- For datasets of 1GB to 20GB, we recommend DVD or DAT.
- For datasets of less than 1GB, we recommend CD or DVD.

1.3 Collection Details

When supplying data, the following are required to allow proper assessment and use of swathe bathymetry;

Vessel Offsets
The X,Y,Z offsets between each sensor and the vessel reference point (RP) should be measured, recorded and input to the navigation computer before the start of survey. The RP is conventionally placed either at the approximate centre of gravity (CoG) of the vessel or at the motion sensor.

Positioning
Before commencing survey work the survey positioning system should be checked for accuracy using known survey points on land. This should ideally be done before the multibeam calibration. All position data must be logged during the survey. Usually this is in a combined file with the raw multibeam data.

Motion Sensor Data
During installation the motion sensor should be corrected for alignment errors in pitch and roll. This can either be done using land survey techniques while the vessel is dry docked, or the sensor can be “zeroed out” while the vessel is at a tight mooring alongside. The correction values should be recorded and entered into the appropriate software. This should be done before the multibeam system is calibrated. Whilst surveying all motion sensor data must be logged. Usually this is in a combined file with raw multibeam data.

Heading Data
During installation the heading sensor should be checked for gross errors and any alignment errors (the difference between the heading of the sensor and the heading of the vessel). These can be checked while the vessel is alongside at a tight mooring using land survey techniques. The correction value should be recorded and entered into the appropriate software. Whilst surveying all heading data must be logged. Usually this is in a combined file with raw multibeam data.

Multibeam Data
Before commencing a survey the multibeam system must be calibrated for timing and alignment errors (latency, pitch, roll & yaw) by carrying out a patch test (see Annex A). The results from the patch test should be recorded and entered into the navigation software. During survey all the raw multibeam data must be logged. Usually this is in a combined file with raw position, motion sensor and heading sensor data.

Sound Velocity in Sea water (SV)
The speed of sound in sea water (SV) must be measured in the survey area at intervals dependent on the environment being surveyed. At a bare minimum the SV should be measured at least once every day. But if working in an environment with a suspected variable SV (such as near the mouth of a river) then more frequent measurements should be taken, preferably at least every 4 hours or if the sound velocity is deemed to have changed by 1 m/s or more. This data must be entered in to the navigation software to allow for online refraction correction of the multibeam system. Depending on the navigation software used these files can be embedded in the raw multibeam logged files or logged as a separate file.

Tidal Data
Tidal information is required at the survey site. Preferably this is measured by one or more tide gauges (depending on the size of the area) deployed in the survey area for the duration of the survey. If this is not possible then a gauge at a nearby port could be used and the tidal data extrapolated using co-tidal information. Tide height should be recorded at 10 minute intervals with the high and low waters also being recorded. Some gauges can self compensate for air pressure changes, but if using a bottom mounted pressure gauge, air pressure readings must be recorded at the same interval as the tide gauge. This can be used when processing the tide gauge data to compensate for air pressure changes. Care must be taken as to the units of measurement such as atmospheres or kPa.

If a hard copy plot and/or .dwg drawing is available, please also supply that, as it helps UKHO to check that the digital data is complete. Drawings may also include useful non-bathymetric detail, such as coastline and details of construction projects. Please record any differences between the digital data and the drawing, especially if depths on the drawing have been edited, but the edits have not been applied to the digital data.

2.0 Value Added Service

When processing and quality controlling data, hydrographic and bathymetric data centres of the ICES community shall strive to meet the following guidelines

2.1 Quality Control

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 data supplier. Other experts in the field or other Data Centres may also be consulted.

Ideally, the accuracy of both the depth and position of soundings should meet IHO S44 Order 1 requirements (see references) and the density of the data should be such that the minimum target sizes stated in IHO S44 Order 1 are always detected. A report should describe how the data was gathered and processed. It is particularly useful if the report can include:

- A list of the equipment and software used, and the personnel involved.
- How positioning equipment was set up and used
- How the echo-sounder transducer was set up and levelled
- How tides were measured, how the tide pole or tide gauge was levelled and how depths were reduced to chart datum. NB a levelling diagram is a useful method of confirming that datum adjustments have been carried out correctly.

Error Budget
If a fully developed error budget based on the specifications and experience of the system has been developed, this will also be extremely useful in the assessment of the data. See Reference - IHO Standards for Hydrographic Surveys.

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 Principal Investigator (PI) or data supplier. Other experts in the field or other Data Centres may also be consulted

2.3 History Documentation

Report of Survey (ROS)

If a report of survey has been produced (regardless of format), this will also be extremely useful in the assessment of the data. A template ROS can be supplied from UKHO if requested – see Annex B.

3.0 Providing Data and Information Products

When addressing a request for information and/or data from the User Community, the hydrographic and bathymetric 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

When delivering Multibeam Survey data the following items are required. :

Vessel Offsets: The X, Y, Z offsets between each sensor and the vessel reference point should be supplied.

Geodetic / Tidal Information: The datum, spheroid and projection used for the survey should be provided. The tidal datum should be described, along with how this ties into other datums (i.e. levelled to UK Ordnance Datum). Any levelling diagrams and benchmark information is useful.

Notable Features: Comments should be provided if at all possible on any notable features in the data, e.g.:

- shoal depth values that may represent buoys, shoals of fish, side echoes off quay walls, electrical noise, etc.
- details of wreck investigations.
- charted dangers not found during the survey.

Alignment Error Data: All recorded values for the motion and heading sensor alignment errors should be provided along with the values obtained from the multibeam patch test.

Processed Data Files: All processed data files should be supplied along with the relevent sound velocity files and processed tide files. Information outlining the directory structure, number and size of files should be supplied to ensure that no data is missing. The data should not be gridded in any way – the actual soundings should be supplied.

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, see below)
- Any problems encountered with the data and their resolution
- Any changes made to the data and the date of the change

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.

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.

3.4 Data to be submitted

The guidance above is to be followed wherever possible. However, if all the above requirements cannot be met this should not prevent the data from being submitted. All data collected is potentially of use especially in areas where existing data is sparse and/or old.

References

IHO Standards for Hydrographic Surveys 4th Edition April 1998 Special Publications No 44
http://www.iho.shom.fr/publicat/free/files/S-44-eng.pdf

Annex A - Suggested Patch Test Procedure

Latency Test - (Sharp Target in Shallow Water). Produce two overlaid longitudinal profiles (of different colours) clearly showing the target at nadir after a latency correction has been determined and applied. The two profiles should be obtained by performing two co-linear and co-directional survey lines (one at slow speed and one at high speed) directly over a sharp target. The report should briefly detail the method and result of the calibration and should clearly state that the observations were undertaken with all MBES latency and attitude corrections set to zero.

Pitch Test – (Sharp Target in Deep Water). Produce two overlaid longitudinal profiles (of different colours) clearly showing the target at nadir after both pitch and latency corrections have been determined and applied. The two profiles should be obtained by performing two co-linear survey lines (both at standard survey speed) on reciprocal headings directly over a sharp target. The report should briefly detail the method and result of the calibration and should clearly state that the observations were undertaken with all MBES attitude corrections set to zero, but with the deduced latency correction applied.

Roll Test – (Flat seabed in Deep Water). Produce two cross-profiles (of different colours) clearly showing the same section of seabed after pitch, latency and roll corrections have been determined and applied. The two profiles should be obtained by performing two co-linear survey lines (both at standard survey speed) on reciprocal headings over a flat section of seabed. The report should briefly detail the method and result of the calibration and should clearly state that the observations were undertaken with the deduced latency and pitch corrections applied but with all other MBES attitude corrections set to zero.

Yaw Test – (Sharp target in Deep Water). Produce two co-registered contour plots (of different colours) clearly showing the target at a large off-track distance after latency, pitch, roll and yaw corrections have been determined and applied. The two data sets should be obtained by performing two parallel, co-directional survey lines with 25% overlap. The first line should have the target close to the starboard extremity of the swath. The second line should have the target close to the port extremity of the swath. The report should briefly detail the method and result of the calibration and should clearly state that the observations were undertaken with the deduced latency, pitch and roll corrections applied but with all other MBES attitude corrections set to zero.

Pitch/Roll Cross-Correlation Test (Wobble Test) – (Flat Seabed in Deep Water). Sail the vessel at survey speed in such a direction relative to the sea and swell so that pitch and roll are maximised. Adjust the IMU misalignment value (in the motion sensor software) around the Z-axis so as to subjectively minimise any degradation (wobble) on the outer beams (this wobble is due to pitch/roll cross-correlation). The report should briefly detail the method and result of the calibration and should clearly state that observations were undertaken with the deduced latency, pitch, roll and yaw corrections applied.

Annex B – Contact Details – UK Hydrographic Office

Digital survey data (and associated information) should be addressed to:

Seabed Data Centre
United Kingdom Hydrographic Office
Admiralty Way
Taunton
Somerset
TA1 2DN
United Kingdom

ICES Guidelines for Surface Drifting Buoy Data

(Compiled March 2001, revised August 2006)

Drifting buoys (UNESCO, 1988) have a long history of use in oceanography, starting in late 1978 with the First GARP Global Experiment (FGGE), principally for the measurement of currents by following the motions of floats attached to some form of sea anchor or drogue. Since 1988, over 1500 Lagrangian drifters have been deployed in the world oceans in the Surface Velocity Program (SVP) of the World Ocean Circulation Experiment (WOCE) and the Tropical Ocean and Global Atmosphere Program (TOGA). The buoys were standardised in 1991, with small spherical hull and floats, and large Holey-Sock drogue centred at 15 meters below the surface. Since 1993, Lagrangian Drifters with barometer ports and other sensors, including thermistor chains, have been in operation. Separate guidelines are available for Profiling Floats that record sub-surface variables.

1.0 Receiving Data

Data Centres may receive these data by several mechanisms:

- Raw data from the transmission service (e.g. Service Argos, Iridium)
- Real-time messages via the Global Telecommunication System (GTS)
- Delayed-mode data from the Principal Investigators (PI).

The Data Centres require the following information to be supplied by the data supplier together with the data.

1.1 Data Standard

Delayed-mode quality controlled drifting buoy data provided by PIs to the Data Centres should contain:

- A full description of the data format used for the data submission.
- Metadata information about the calibration (equations and coefficients) applied to the data set.
- Quality controlled data, reported at the original sampling interval of the instrumentation.
- A description of the quality control procedures applied to the data set.

All observed variables should be clearly specified and described. If parameter codes are used, then the source data dictionary must be specified in the metadata documentation. Variable units and precision must be clearly stated. If computed values are included, the equations used in the computations should be stated.

All relevant calibrations should be applied to the observed 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.

1.2 Format Description

Data Centres may receive drifting buoy messages in real-time coded formats transmitted on the GTS and in delayed-mode quality controlled drifting buoy data formats from PIs. The WMO coded formats (WMO, 1995) used for GTS distribution of real-time drifting buoy messages are WMO FM 18-XII BUOY (ASCII), FM 94 BUFR (binary).

The contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange (http://www.ices.dk/ocean/formats/getade_guide.htm) prepared by IODE.

The PIs data formats for the exchange of delayed-mode surface drifting buoy data set should include:

- A full description of the format used (preferably a fully documented ASCII format).
- The files should be homogeneous (i.e. each piece of information must always be in the same place in the file).
- Individual fields and units should be clearly defined.
- Ideally all of the data from the instrument should be stored in a single file.
- All data values should be in SI units.
- Time reported in UTC is strongly recommended.

1.3 Collection Details

Details on the collection of drifting buoy data should always be included. Metadata such as WMO number, buoy type, drogue type, etc., are important for proper use and understanding of the data. While the BUOY format can allow some extra data in real-time, BUFR can provide more. Refer to the drifting buoy BUFR template listed in 1.2 for the list of variables that should be included.

Metadata requirements for delayed-mode quality controlled drifting buoy data provided by PIs to the Data Centres include:

- Deployment platform name
- Country, organisation, Principal Investigator
- Project name
- Float number
- WMO number
- Sensor resolutions
- Information about the instruments and sensors (type and manufacturer, serial and model numbers, board type and serial number, software version)
- 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 should also be included.

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

The primary responsibility for data quality control lies with the PIs from which the observations originate. In many cases, drifting buoy data originate from a national meteorological service, an oceanographic institute or a PI for a particular research project.

Real-time

All real-time drifting buoy data go through Service Argos to be distributed on to the GTS. PIs and other data centres and services carry out appropriate quality control of observational messages they receive. Automatic Quality Control (QC) checks can quickly detect erroneous messages by comparing sensor data with constant limits.

Variable	Lower limit	Upper limit
Sea level air pressure (hPa)	800	1080
Station air pressure (hPa)	400	1080
Air pressure tendency (hPa/3H)	0	100
Water temperature (deg)	- 1.8	+ 45
Air temperature (deg)	- 80	+ 50
Wind speed (m/s)	0	100
Wind direction (deg)	0	360

The DBCP QC Guidelines for GTS Buoy Data (http://www.dbcp.noaa.gov/dbcp/2qgd.html) was set up to exchange QC information amongst participants. Erroneous messages and questionable data can be brought to the attention of buoy operators, manufacturers and others interested in the quality of GTS drifting buoy data by posting to the distribution list buoy-qir@vedur.is.

Delayed-mode

There are three main components to the quality control of delayed-mode surface drifting buoy data. All three components are used at the Data Centres or the PI to quality control the datasets.

The first component examines the characteristics of each float track looking to identify errors in either position or time based on calculated buoy speed versus time.

The second component is subjective as each observed variable is viewed independently to identify values that appear to be outside prescribed climatic limits for the area in question. A time series of one month of observations for each individual buoy is treated at one time. Knowledge of the different types of real and erroneous oceanographic and meteorological features is critical. This knowledge, when combined with a local knowledge of water mass structure, statistics of data anomalies, and cross validation with climatological data, ensures a data set of the best possible quality.

The third component is to identify and eliminate duplicate data. Duplicate observations 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.

To deal with both the real-time and delayed mode data, it is recommended that the Data Centre manage surface drifting buoy data in a continuously managed database. This will provide to the client those messages reported in real-time when these represent the only version available, or the delayed-mode data of higher quality which replace the original real-time data set. An overview of the data management practices for delayed mode drifting buoy data in a continuously managed database is provided in Annex A (Wilson, 1998).

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. The Data Centre may also consult the Joint IOC/WMO Data Buoy Cooperation Panel (DBCP)(http://dbcp.noaa.gov) or ISDM (Formerly MEDS), the IODE Data Center for Drifting Buoys for advice when needed.

2.3 History Documentation

All procedures applied to a dataset should be fully documented by the Data Centre. These include all quality control tests applied and should accompany that dataset. All problems and resulting resolutions should 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 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
- Variable 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. buoy type, drogue type, 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 in Section 2.1)
- Any problems encountered with the data and their resolution
- Any changes made to the data and the date of the change

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.

3.3 Referral Service

ICES member research and operational data centres produce a variety of data analysis products and referral services. Data and information products are disseminated as widely as possible and via a number of media including mail, electronic mail, internet websites, ftp and CD/DVD. The DBCP supports several Action Groups who produce data and information for specific ocean areas. 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.

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

UNESCO. 1988. Guide to Drifting Data Buoys, IOC/WMO Manuals and Guides # 20.

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

WMO. 1995. WMO Manual on Codes No. 306.

UNESCO. 1991. Manual on International Oceanographic Data Exchange, IOC/ICSU Manual and Guides # 9, Revised Edition.

Annex A

Both the real-time messages and delayed mode surface drifting buoy data are available at the sample interval set by the manufacturer. This interval is typically hourly or synoptic hours depending on the requirements. The delayed mode data undergo calibration and quality control often incorporating site specific knowledge and experience of the PI. Real-time messages are often those surface drifting buoy data that undergo automatic, bulk quality control tests within operational time frames. The extensive quality control incorporating site-specific knowledge and experience of the PI often take longer. Real-time messages are most useful to those involved in operational forecasts, while delayed mode data are more useful to research.

To manage surface drifting buoy 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 messages 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.

ICES Guidelines for Biological Plankton Data

(Compiled August 2001; reviewed April 2006)

In the context of this guideline, phytoplankton or zooplankton sampling may be accomplished using either a vertical, horizontal or oblique tow of a net or from a rosette bottle.

In the case of a net, such a device would consist of frame which houses a mesh used in collecting the sample. An example maybe a square frame with multiple nets, of a single, conical shaped mesh with a circular ring opening. Typical mesh sizes would be less than 1000 μm (microns). At the mouth end the opening may be of up to 2 m. Attached at the small end of the net would be a jar or cod-end with a typical opening of about 10 cm.

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.

1.1 Data standard

The data set should consist of header and data information in one or more standard ASCII files. Each record should consist of date and time, navigation data, and measured parameters. It is recommended that each cruise constitute a single file. The navigation data should be in ASCII and in the form of latitude and longitude in degrees and decimal minutes, or decimal degrees. (Explicitly state which format is being used. It is recommended that N, S, E and W labels are used instead of plus and minus signs). Date and time must include month, day, year, hour, and minute. It is recommended that UTC be used.

All parameters must be clearly specified and described. If parameter codes are to be used, then the source data dictionary must be specified. 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 data including laboratory and field calibrations. 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.

Sufficient self-explanatory information and documentation should accompany the 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.

A brief description of the sample and data processing procedures must be included and should contain information regarding:

- Laboratory procedures and instrumentation
- Any species counts or mass measurements
- Description of any respiration, feeding or physiological experiments and results (e.g. carbon dioxide rates, carbon and nitrogen measurements)
- Report on corrections, editing or quality control procedures applied to the data
- Time reported in UTC is strongly recommended
- Estimate of final uncertainty in the data

Information about any supplementary/complementary data collected at the same time should also be supplied.

If a cruise/data report is available describing the data collection and processing, this can be referenced. If possible a copy should be supplied with the data.

1.2 Format Description

Data should be supplied in a fully documented ASCII format. Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is strongly recommended. The contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange (http://www.ices.dk/ocean/formats/getade_guide.htm) 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

Other pertinent information to be included in the data transfer to the Data Centre includes:

- Project, ship, cruise identifier
- Country, organisation
- Date, time, latitude and longitude (for start and end if sampling via a net tow)
- Sounding, maximum and minimum pressure or depth of the tow
- Description of operational procedures such as tow orientation (vertical, horizontal or oblique), methods of position fixing (e.g. DGPS, GPS, etc.)
- Weather conditions (including sun and wind)
- Gear type (e.g. net mesh size, net mouth size, single or multi-net, etc.)
- Sample preservation method (e.g. pickling, frozen, etc.)
- Sample analysis/processing or data collection procedures (e.g. filtered size ranges, sub-sampling, etc.)

Any additional information of use to secondary users which may have affected the data or have a bearing on its subsequent use. An example field log sheet is included in Annex A.

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

A range of checks are carried out on the data to ensure that they have been imported into the Data Centre's format without any loss of information. These checks should include:

- General check of accompanying information (e.g. tow dates within cruise dates, correct cruise identifier)
- Plot navigation to ensure no land points; compare with cruise report/CSR track chart if available
- Flag suspicious data or correct after consultation with the data supplier
- Checks on ship speed

If the navigation data are supplied separately, they will be merged with the individual tows.

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 Principal Investigator (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. This is described in more detail below:

- 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 data file (e.g. ship, cruise, project, net tow deployment identifiers, start and end dates and times of tows, 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
- 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.

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.

Annex A
Example Net Tow Log Sheet

General

Project:________________ Ship:___________________ Country:_____________

Cruise Number:_________ Tow Number:____________ Event Number:________

Location:____

Bottom Sounding:________________ Weather:_____________________________________

Wind:_______________________________________

Start Tow

Date:___________________________ Time (UTC):______________ Twilight:_____

Latitude:________________________ Longitude:________________ Method:_____

End Tow

Date:___________________________ Time (UTC):______________ Twilight:_____

Latitude:________________________ Longitude:________________ Method:_____

Net Mouth Size:______________ Wire Angle:________________

Net	Depth Range (Wire out)	Mesh Size	Volume of Water Filtered

Comments:

ICES Guidelines for Moored ADCP Data

(Compiled October 1999, revised August 2001; August 2006)

ADCPs (acoustic doppler current profiler) were first introduced to the oceanography community in the late 1970s (Rowe and Young, 1979). The instrument measures water velocity over a range of depths using doppler shifts in active acoustic signals. ADCPs may be moored on a traditional oceanographic mooring (these guidelines apply), mounted on a bottom-frame or attached to the bottom of a ship (see shipboard ADCP guidelines). Different data management requirements exist for the three configurations.

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.

1.1 Data standard

Raw and quality controlled ADCP data should, whenever possible, be stored at the original sampling frequency. The ADCP data should be accompanied by an instrument record/deployment sheet which should include such information as described under Collection Details (Section 1.3).All parameters must be clearly specified and described. If parameter codes are to be used, then the source data dictionary must be specified. 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 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. or profiles. An explicit statement should be made of the checks and edits applied to the data. If any data values have been removed, the times of the removed values should be noted. If the instrument shows differences in measured values when compared to traditional instruments (e.g. Aanderaa Instruments RCM X instruments) it should be sent to the manufacturer for calibration.

Sufficient self-explanatory information and documentation should accompany the 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.

A brief description of the data calibration, quality and processing must be included and should contain information regarding:

- Laboratory calibrations
- Manufactures calibrations
- In-situ calibrations (e.g. using a calibrated CTD to supply temperature data to calibrate any temperature sensor that may be installed on the ADCP or a current meter moored on the same ADCP mooring to provide water current data for comparison tests.)
- The time zone in use should be clearly stated and each data cycle should include date/time of observation (without loss of precision). It is recommended that UTC is used.
- Estimate of final uncertainty in the data

A brief description of the data processing procedures (manufacturers and in-house) must be included and should contain information regarding:

- Filtering, de-spiking, or smoothing methods
- Editing or quality control methods
- Interpolation techniques
- Adjustments made due to variations in calibration

If a cruise/data report is available describing the data collection and processing, this can be referenced. If possible a copy should be supplied with the data.

1.2 Format Description

Data should be supplied in a fully documented ASCII format. Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is strongly recommended. Ideally all of the data from the instrument should be stored in a single file. The contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange (http://www.ices.dk/ocean/formats/getade_guide.htm) 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

Other pertinent information to be included in the data transfer to the Data Centre includes:

- Name of the country and organisation responsible for the deployment, recovery, collection and processing of the data. The name of the Principal Investigator should be included.
- Project, ship, mooring type, mooring number
- Dates and times of each instrument deployment and recovery
- Dates and times of start and end of usable data for each instrument
- Details of the instrument and sensors (e.g. manufacturer, instrument type, model number, serial number and any modifications carried out, number of transducers)
- Description of operational procedures including sampling interval (time between ensembles), pings per ensemble, bin size, number of bins, percentage good level, automated data rejection (e.g. fish rejection algorithms), etc.
- Frequency (kHz), band type (broad, narrow), heads facing upward or downward.
- Latitude and longitude, method of position fix (e.g. GPS, DGPS)
- Parameters collected (e.g. u and v velocity components, vertical velocity, error velocity, echo intensity, percent good pings)
- Multiplicative and/or additive scale factors where applicable
- Water column depth (specify method e.g. sounding and methodology, chart, etc.)
- Instrument depth (or height, specify which)

See the attached **ADCP Instrument Sheet** as an example of what could be used at sea and what could be sent to the Data Centre.

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

A range of checks are carried out on the data to ensure that they have been correctly imported into the Data Centre's format without any loss of information. For self-contained ADCP data, these should include:

- General check of accompanying information
- Automatic range checking of each parameter
- Visual inspection of the time series (e.g. time series plot, current vector scatter plot, progressive vector diagram, etc.)
- Flag spikes in the data
- Flag suspicious data or correct the data after consultation with the data supplier
- Check against other data collected on nearby moorings
- Check corrections/calibrations applied
- Check latitude/longitude not on land

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 Principal

Investigator (PI) or data supplier. Other experts in the field or other Data Centres may also be consulted. Problems solved shall be reported to the originating Principal Investigator (PI) or data supplier.

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 data set should accompany that data set. 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. This is described in more detail below:

- A data format description fully detailing the format in which the data will be supplied
- Any ancillary parameters (e.g. temperature)
- Parameter and unit definitions and scales of reference
- Definition of flagging scheme, if flags are used
- Relevant information included in the data file (e.g. geographical position – latitude, longitude and water depth, self-contained ADCP deployment identifiers, start and end times of data, sampling interval, project name, 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
- Any changes made to the data and the date of the change

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.

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

Rowe, F. and J. Young. 1979. An ocean current profiler using doppler sonar, Oceans '79 Proceedings.

INSTRUMENT SHEET (ADCP)

Record No		**Acoustic Frequency: kHz**
Instrument Type	ADCP	**Broadband / Narrowband**
Instrument Serial No.		**Upward / Downward**

Deployment Position and type (e.g. GPS, DGPS)			
Area/Project			
Sounding (metres)		**Inst. Heads height above sea floor (m)**	

Instrument ON:		**Instrument OFF:**	
Date:		**Date:**	
UTC Time:		**UTC Time:**	
IN Water:		**OUT Water:**	
Date:		**Date:**	
UTC Time:		**UTC Time:**	

Instrument Configuration			
Pings per ensemble		***Parameters Recorded***	
Bin Length		**Velocity**	Y / N
Number of Bins		**Co-ordinate system:** **Earth/ Beam/ Instrument/ Ship**	E / B / I / S
Blank after Transmit		**Correlation**	Y / N
Profile Mode		**Intensity**	Y / N
Time between ping groups		**Percentage Good Pings**	Y / N
Time per Ensemble		**Record Status? (reason ADCP rejects data)**	Y / N
Deployment Length		**Recorded Enabled**	Y / N

Battery Capacity Fitted		**MB Memory Available**	
Samples Expected		**Samples Received**	
Pre. Mooring Checks		**Post Recovery Check**	

Test and Configuration Files:

Readings and Remarks:

ICES Guidelines for Shipboard ADCP Data

(Compiled May 1999; revised April 2006)

ADCPs (acoustic doppler current profiler) were first introduced to the oceanography community in the late 1970s (Rowe and Young, 1979). The instrument measures water velocity over a range of depths using doppler shifts in active acoustic signals. ADCPs may be moored on a traditional oceanographic mooring (see moored ADCP guidelines) or attached to the bottom of a ship (these guidelines apply). Different data management requirements exist for the two configurations.

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.

1.1 Data Standard

Data should be provided as time series of profiles. Each profile should contain information describing the profile date, time, and position. Time reported in UTC is strongly recommended. Each profile should contain numerous records of data, where each record represents a vertical bin in the profile. For each bin, there should be the centre depth of the bin, three orthogonal velocity components, percent good pings, the error velocity, as well as average echo amplitude and average beam correlation (the average of the 4 beams).

All parameters must be clearly specified and described. If parameter codes are to be used, then the source data dictionary must be specified. 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 data including laboratory and field calibrations. The data should be fully checked for quality and flagged for erroneous values such as spikes, gaps, or profiles. An explicit statement should be made of the checks and edits applied to the data.

Sufficient self-explanatory information and documentation should accompany the 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.

A brief description of the data calibration, quality and processing must be included and should contain information regarding:

- Reported velocities should be in positive North (v) and positive East direction (u). If not, the co-ordinate system of the horizontal velocity components must be clearly stated.
- The criteria used for flagging or rejecting data (for example threshold values of error velocity and /or percent good pings) must be discussed.
- The method used for correcting profiles for ship speed (that is bottom tracking, navigation or reference level of no motion) must be described.
- Calibrations carried out to correct for transducer misalignment.
- Any problems of contamination of the data due to bubbles in rough weather, high ship speed (propeller noise), change in ship direction or ship speed zero and how these are dealt with.
- An estimate of final accuracy in the data.
- Gyro corrections made utilising heading information from differential GPS systems along with details of the DGPS equipment.
- The method of calculating sound speed should be stated. If a constant value is employed, state what it is and how it might vary along the cruise track. If temperature and a fixed salinity were used to estimate sound speed, state what the latter value is and where temperature data are stored.
- There should be specific mention of bottom tracked velocities. These are valuable in the shallow sections of the cruise for calibration information.
- Comments on Time drift of the ADCP clock
- Number of transducers

- Parameters collected (e.g. u and v velocity components, vertical velocity, error velocity, echo intensity, percent good pings)

If a cruise/data report is available describing the data collection and processing, this can be referenced. If possible, a copy should be supplied with the data.

1.2 Format Description

Data should be supplied in a fully documented format (ASCII, NetCDF, …). Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is strongly recommended. Ideally all of the data from the instrument should be stored in a single file. The contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange (http://www.ices.dk/ocean/formats/getade_guide.htm) 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

Other pertinent information to be included in the data transfer to the Data Centre includes:

- Project, ship, cruise identifier
- Country, organisation
- Details of the instrument and sensors (e.g. manufacturer, instrument type, model number, serial number and any modifications carried out, number of transducers)
- Description of operational procedures including sampling interval (time between ensembles), pings per ensemble, bin size, number of bins, bottom tracking on/off, pitch and roll on/off, percentage good level, method of position fix (e.g. GPS, DGPS), automated data rejection (e.g. fish rejection algorithms), etc.
- Frequency (kHz), band type (broad, narrow)
- Date and time of the start and end of the profiles for each data file

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

A range of checks are carried out on the data to ensure that they have been imported into the Data Centre’s format without any loss of information. For self-contained ADCP data, these should include:

- General check of accompanying information
- Automatic range checking of each parameter
- Visual inspection of the time series
- Flag spikes in the data
- Flag suspicious data or correct the data after consultation with the data supplier
- Check corrections/calibrations applied
- Plot navigation data to ensure no land points; compare with cruise report/CSR track chart if available
- Compare the data with available bathymetry (if no bottom ping is used) and flag the “under bottom” data

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 Principal Investigator (PI) or data supplier. The Data Centre may also consult the Japan Oceanographic Data Centre, as the IOC Responsible National Oceanographic Data Centre for shipboard ADCP data or the University of Hawaii ADCP centre. 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. This is described in more detail below:

- A data format description fully detailing the format in which the data will be supplied
- Any ancillary parameters (e.g. temperature)
- Parameter and unit definitions, and scales of reference
- Definition of flagging scheme, if flags are used
- Relevant information included in the data file (e.g. ship, cruise, project, start and end dates, 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
- Any changes made to the data and the date of the change

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.

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

Joint Archive for Shipboard ADCP at http://ilikai.soest.hawaii.edu/sadcp/

Joint Archive for Shipboard ADCP at http://ilikai.soest.hawaii.edu/sadcp/

Rowe, F. and J. Young. 1979. An ocean current profiler using doppler sonar, Oceans '79 Proceedings.

ICES Guidelines for SeaSoar (Batfish) Data

(Compiled May 1999, revised August 2001, revised December 2006)

The Seasoar (or Batfish) instrument is a towed platform typically equipped with a CTD. Other sensors, such as fluorometers or transmissometers, may also be added. The SeaSoar has hydraulically controlled wings, which give it the ability to "fly" up and down in the water column. On a typical deployment, the SeaSoar is towed behind a ship at speeds of up to 10 knots using a faired cable while flying between two depths, Use of a faired cable reduces the drag on the cable and allows the SeaSoar to fly sawtooth profiles with a maximum depth of ~500m. Without fairing or with damaged fairing the depth capabilities of the SeaSoar are dramatically reduced. With no fairing at all the SeaSoar would struggle to attain 100m depth.

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.

1.1 Data standard

Data should be provided as time series (of 1-second averages) and should be merged with navigation data. If the navigation has not been merged, they should be submitted as a separate file, which could be linked to the SeaSoar data using date and time.

If data are not available in this form, then data split into 'pseudo-CTD' casts are acceptable. In this context, a 'pseudo-CTD' cast is defined as that part of the SeaSoar profile collected on the downward undulation of the instrument. It is also acceptable to provide data in a gridded form, for example, eight decibar vertical resolution and the horizontal resolution to match the wavelength of the SeaSoar oscillation through the water. The horizontal resolution of the gridding should allow for inclusion of two full oscillations of the SeaSoar through the water column. The method used for the generation of ‘pseudo-CTD’s’ should be fully described.

All parameters must be clearly specified and described. If parameter codes are to be used, then the source data dictionary must be specified. 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 data including laboratory and field calibrations. The data should be fully checked for quality and erroneous values such as spikes or gaps should be replaced with a suitable ‘No data’ value. An explicit statement should be made of the checks and edits applied to the data.

Sufficient self-explanatory information and documentation should accompany the 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.

A brief description of the data calibration, quality and processing must be included and should contain information regarding:

- Laboratory calibrations (e.g. whether carried out in accordance with SCOR Working Group 51 recommendations. See also Karl (1996) and UNESCO (1991).)
- In-situ calibrations (e.g. lowering a separate, calibrated CTD before and after a SeaSoar deployment and then towing SeaSoar through CTD station, use of thermosalinograph, or water samples taken from the non-toxic supply)
- Report on corrections made to data especially for offsets in salinity due to fouling of the conductivity cell
- Time reported in UTC is essential
- Estimate of final uncertainty in the data

A brief description of the data processing procedures (manufacturers and in-house) must be included and should contain information regarding:

- Instrument Details

- Filtering, de-spiking or smoothing methods
- Editing or quality control methods
- Time lag correction scheme for temperature sensor
- Adjustments made due to variations in calibration

If a cruise/data report is available describing the data collection and processing, this can be referenced. If possible a copy should be supplied with the data.

1.2 Format Description

Data should be supplied in a fully documented ASCII format. Individual fields, units, etc. should be clearly defined and time reported in UTC. The contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange (http://www.ices.dk/ocean/formats/getade_guide.htm) 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

Other pertinent information to be included in the data transfer to the Data Centre includes:

- Project, ship, cruise identifier, SeaSoar deployment identifiers
- Country, organisation
- Date and time of the start and end of the SeaSoar run
- For data supplied as 'pseudo-CTD' casts; date, time, latitude, longitude, and an up/down cast indicator for each cast
- Details of the instrument and sensors (e.g. manufacturer, model number and any modifications carried out)
- Description of operational procedures including sampling rate, sensor resolutions, undulation rate, methods of position fixing (e.g. DGPS, GPS, etc.)

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

A range of checks are carried out on the data to ensure that they have been imported into the Data Centre’s format without any loss of information. For SeaSoar data, these should include:

- General check of accompanying information (e.g. SeaSoar runs within cruise dates, correct cruise identifier)
- Automatic range checking of each parameter
- Plot navigation file to ensure no land points, no spikes and sensible vessel speed; compare with cruise report/CSR track chart if available
- Visual inspection of time series (or ‘pseudo-casts’) for all parameters supplied - this may include oxygen, transmittance or fluorescence, in addition to temperature and salinity
- Removal of spikes in data (or gridding artefacts), replacing with defined ‘No Data’ value.
- Flag suspicious data or correct after consultation with the data supplier
- Check corrections/calibration applied
- Compare with thermosalinograph data and calibration CTD casts if available
- Check available comparison with water bottle samples, and corrections/calibrations applied
- Compare with climatology

If the navigation data are supplied separately, they will be merged with the SeaSoar data at the Data Centre.

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 Principal Investigator (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. This is described in more detail below:

- 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 data file (e.g. ship, cruise, project, SeaSoar deployment identifiers, start and end dates and times of SeaSoar run, 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
- Any changes made to the data and the date of the change

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.

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

Karl, David, Luis Tupas, Fernando Santiago-Mandujanu, Craig Nosse, Dale Hebel, Eric Firing and Roger Lukas. 1996. Hawaii Ocean Time-Series Data Report 7:1995, SOEST 96-09, University of Hawaii.

Pollard, R, 1986, Frontal surveys with a towed profiling conductivity /temperature/depth measurement package (SeaSoar)*Nature* 323, 433 - 435 (02 October 1986)

UNESCO. 1988. The acquisition, calibration and analysis of CTD data. A report of SCOR WG 51. Tech. Pap. Mar Sci., 54: 59pp.

UNESCO, 1991. Processing of Oceanographic Station Data, JPOTS Editorial Panel.

ICES Guidelines for Surface Underway Data

(Compiled May 1999, revised August 2001; April 2006)

Underway near-surface measurements are typically made using the cooling water intake of a vessel. This guideline refers to electronically measured parameters typically including temperature and salinity. Additional parameter measurements such as fluorescence and nutrients may also be added. Autosampling devises may also be included in the underway system, thus providing the automatic water sampling for lab analysis and calibration of the electronic equipment.

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.

1.1 Data standard

Data should be provided as time series of averages and should be merged with navigation data. The typical time interval of the averages should be between 3 minutes (delayed mode transmission) and 30 minutes (real time transmission).except when in frontal regions when the averaging interval should be one minute. Alternately, the entire dataset may be provided in one-minute intervals, rather than mixing the 30 and one-minute intervals.

The format of the data set should be a single, standard ASCII file with one average value per record. A record should consist of date and time, navigation data, observed parameters and number of raw observations in the average. It is recommended that each cruise constitute a single file, but for operational oceanography it is recommended to have one file per day of cruise. The navigation data should be in ASCII and in the form of latitude and longitude in degrees and decimal minutes, or decimal degrees. The navigation format should be explicitly stated. Date and time must include month, day, year, hour, and minute. It is recommended that UTC is used.

If the navigation has not been merged, they should be submitted as a separate file, which could be linked to the along track data using date and time. The navigation position should be applicable to the middle of the averaging interval.

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

The list of parameters may include temperature, salinity, fluorescence, dissolved oxygen, transmittance, nutrients and partial CO_2.

All relevant calibrations should be applied to the data including laboratory and field calibrations.
If no calibration have been applied, this should be clearly stated. 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.

A brief description of the data calibration and quality issues must be included and should contain information regarding:

- Laboratory calibrations for all parameters (e.g. for temperature and salinity see SCOR Working Group 51 recommendations. Although applicable to lowered CTD measurements, there is relevant information in this report. As well, see Karl (1996) and UNESCO (1991).)
- Computation equations used to determine the parameters
- In-situ calibrations (e.g. comparisons with lowered near-surface CTD data, or water samples)
- Report on corrections made to data
- Estimate of final uncertainty in the data

A brief description of the data processing procedures must be included and should contain information regarding:

- Filtering, de-spiking or smoothing methods

- Editing or quality control methods
- Time lag correction scheme
- Adjustments made due to variations in calibration

If a cruise/data report is available describing the data collection and processing, this can be referenced. If possible, a copy should be supplied with the data.

1.2 Format Description

Data should be supplied in a fully documented ASCII format. Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is strongly recommended. Ideally all of the data from the instrument should be stored in a single file. The contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange (http://www.ices.dk/ocean/formats/getade_guide.htm) 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

Other pertinent information to be included in the data transfer to the Data Centre includes:

- Project, ship, cruise identifier
- Country, organisation
- Date and time of the start and end of the sampling
- details of the instrument and sensors (e.g. manufacturer, model number, serial number and any modifications carried out, and sampling rate)
- Description of operational procedures including sampling rate, sensor resolutions, methods of position fixing (e.g. GPS, DGPS)
- Flow volume per unit time, flow rate, size of water line
- Depth and position of water intake
- Discussion of possible temperature increase due to the flow through the ships intake lines. An estimate of this increase and whether or not the data (both temperature and salinity) have been corrected for this increase. In addition, how this affects the final uncertainty in the data.
- The location of the sensors should be clearly stated (eg. if temperature sensor is next to the conductivity sensor or mounted separately on the hull). The time lag for a parcel of water moving between the sensors should be stated.

Any additional information of use to secondary users which may have affected the data or have a bearing on its subsequent use should be provided. In particular, if water samples were also collected the following related information should be included:

- The water sample data should be included with the alongtrack data
- Was it an automatic or manual water sampling system, and how were the samples collected, what was the sample volume
- Details of any automatic system used (e.g. manufacturer, model number, serial number)
- What was the water sampling time interval
- What analyses were conducted on the water samples, how much time elapsed between collection and the analysis

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

A range of checks are carried out on the data to ensure that they have been imported into the Data Centre's format correctly and without any loss of information. For alongtrack data, these should include:

- General check of accompanying information (e.g. Alongtrack runs within cruise dates, correct cruise identifier)
- Plot navigation file to ensure no land points; compare with cruise report/CSR track chart if available

- Checks on ship speed
- Automatic range checking of each parameter
- Visual inspection of time series for all parameters supplied - this may include, for example, oxygen, transmittance or fluorescence, in addition to temperature and salinity
- Flag spikes in data
- Flag suspicious data or correct after consultation with the data supplier
- Check any available comparison with water bottle samples, and corrections/calibrations applied
- Compare with climatology

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 Principal Investigator (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. This is described in more detail below:

- A data format description fully detailing the format in which the data will be supplied
- Any ancillary parameters
- Parameter and unit definitions, and scales of reference
- Definition of flagging scheme, if flags are used
- Relevant information included in the data file (e.g. ship, cruise, project, start and end dates, 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
- Any changes made to the data and the date of the change

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.

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

Karl, David, Luis Tupas, Fernando Santiago-Mandujanu, Craig Nosse, Dale Hebel, Eric Firing and Roger Lukas. 1996. Hawaii Ocean Time-Series Data Report 7:1995, SOEST 96-09, University of Hawaii.

UNESCO. 1988. The acquisition, calibration and analysis of CTD data. A report of SCOR WG 51. Tech. Pap. Mar Sci., 54: 59pp.

UNESCO, 1991. Processing of Oceanographic Station Data, JPOTS Editorial Panel.

UNESCO. 2001. IODE Steering Group for Underway Sea Surface Salinity Data Pilot Project, First Session, Brest France, 15-16 November 2001.

SOOP (Ship Of Opportunity Programme) thermosalinograph guide:
http://www.ifremer.fr/ird/soopip/tsg.html#TSG_GUIDE
GOSUD real time QC procedures, version 1.0, 2003, available at:
http://www.coriolis.eu.org/cdc/quality_control.htm

ICES Guidelines for Discrete Water Sample Data

(Compiled December 1999, revised August 2001; June 2006)

In the context of this guideline, discrete water sample data are considered to be any data that result from a single collection of water and so covers a huge variety of parameters. This collection of water must have a specific, identifiable time, position and depth. Such data could originate from a single bottle attached to a rosette or water drawn from a non-toxic supply.

No integrated samples are considered as part of discrete water sample data. Thus, tows that result in integrated data values are not considered in discrete water sample data. Nor are integrated samples from a pumping system or sediment trap.

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.

1.1 Data standard

All parameters must be clearly specified and described. If parameter codes are to be used, then the source data dictionary consistency must be specified. 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.

The data should be fully checked for quality and pre-edited or flagged for erroneous values. An explicit statement should be made of the checks and edits applied to the data.

A brief description, or a reference, to the data collection and processing methods (e.g. reference to a specific technique or specific project protocols) must be included and should contain information regarding:

- Describe or reference full laboratory methods and procedures
- If sample was sent out for analysis, give laboratory name and accreditation level
- Describe or reference any internal or external quality assurance procedures (e.g. QUASIMEME, IAPSO)

A brief description of the data processing procedures must be included and should contain information regarding:

- editing/quality control methods
- how are trace values (values below the detection limit) identified
- how are missing values handled (null vs. zero, or “blanks”)
- what is the precision of the methods (e.g. number of significant figures)
- what analyses has been performed (use parameters descriptions as described in the ICES green book)
- what units are used
- whether any duplicate samples were taken
- describe what quality flags are used if any
- comments describing each station
- supply a calibration document

If a cruise/data report is available describing the data collection and processing, this can be referenced. If possible a copy should be supplied with the data.

1.2 Format description

Data should be supplied in a fully documented ASCII format. Data Centres are capable of handling water sample data in a wide variety of user-defined and project formats. If in doubt about the suitability of any particular format, advice from the Data Centre should be sought.

Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is strongly recommended. Ideally all of the data from the single water source should be stored in a single file. The contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange (http://www.ices.dk/ocean/formats/getade_guide.htm) prepared by the IOC's Group of Experts on the Technical Aspects of Data Exchange (GETADE) and available from RNODC Formats.

Often different groups or laboratories will analyse a single water sample for a multitude of parameters. In such cases, it is common for the data from the different groups to arrive at the data centre at different times. The receiving data centre may merge those data from a single water source. Thus it is crucial that the date/time, position and sample identifier accompany the data.

1.3 Collection Details

Pertinent information to be included in the data transfer to the Data Centre includes:

- Project, platform, cruise identifier
- Country, organisation, institute, PI
- Station number, site details, sample identifier (or bottle number),), type of station (CTD, CTD(NMMP), continuous flow etc.,
- Analyses performed e.g. salinity and nutrients
- Date and time of the start of the sampling and date of analysis (UTC is recommended)
- Position (latitude and longitude degrees and minutes or decimal degrees can be used. Explicitly state which format is being used. It is recommended that N, S, E and W labels are used instead of plus and minus signs.)
- Description of operational procedures including (where applicable) sampling rate, detection limits, standard analytic procedures, calibration of equipment, quality control of original data, methods of position fixing (e.g. GPS, DGPS)
- Details of the collection instrument and sensor (e.g. manufacturer, model number, serial number, and sampling rate)
- Sounding (station depth and sample depth) should be included for each station. The method and assumptions of determining the sounding should be included.
- Type of analyses undertaken including any nutrient samples analysed
- Range of data values (desirable)

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

For additional information on quality control procedures, metadata requirements for particular parameters and collection instrumentation, see UNESCO (1996).

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

A range of checks are carried out on the data to ensure that they have been imported into the Data Centre’s format correctly and without any loss of information. For discrete water sample data, these should include:

- Check header details (vessel, cruise number, station numbers, date/time, latitude/longitude (start and end), instrument number and type, station depth, cast (up and down) data type /no. of data points, platform identifier)
- Plot station positions to check not on land
- Check ship speed between stations to look for incorrect position or date/time
- Automatic range checking of each parameter (e.g. WOD 1998, Maillard 2000)
- Check units of parameters supplied
- Check pressure increasing or decreasing as appropriate
- Check no data points below bottom depth
- Check depths against echo sounder
- Plot profiles (individually, in groups, etc)

- Check for spikes
- Check for vertical stability/inversions
- Check profiles vs. regional climatology
- Check calibration information available
- Compare parameters for predictable relationships (e.g. parameter ratios)
- Check for consecutive constant values
- Duplicate detection when comparing to archived data
- Flag suspicious data or correct after consultation with Principal Investigator (PI)

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 Request for Support

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 to its clients well-defined data or products. If digital data are provided, the Data Centre will provide sufficient self-explanatory series header 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. This is described in more detail below:

- 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 additional quality control
- Flagging scheme, if flags are used
- Data history document (as described in 3.2 below)
- Accompanying data (e.g. CTD data at the time of bottle trip)

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.

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

Maillard, C. And M. Fichaut. 2000. Medar-Medatlas Protocol, Part I : Exchange Format And Quality Checks For Observed Profiles, IFREMER, June 2000 - R.INT.TMSI/IDM/SISMER/SIS00-084.

UNESCO. 1996. IOC-EU-BSH-NOAA-(WDC-A). International Workshop on Oceanographic Biological and Chemical Data Management Hamburg, Germany 20-23 May 1996, IOC Workshop Report 122.

WOD, 1998. World Ocean Database, Documentation and Quality Control, Version 2, Silver Spring, MD, December 1999.