6Th Framework of EC DG Research

SIMORC

DATA QUALITY CONTROL PROCEDURES

Version 3.0
September 2007

Prepared by BODC
Date: September 2007
Website: www.simorc.org

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

1.0 Introduction

The Earth's natural systems are complex environments in which research is difficult and where many unpredictable 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 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:

1. <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.

2. <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.

3. <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.

These are issues which have been addressed by BODC and which have been incorporated in our day-to-day handling of data. This document describes in detail the steps that BODC take to ensure that data provided are of high quality, are easily accessible and are reliable to the extent any natural data can be. However it must be made clear that this document is to be used as a set of *guidelines* for quality control only, as the finer details are often down to human perception and will vary from situation to situation.

2.0 Submitting Data to BODC

2.1 Data Delivery Mechanisms

The following delivery choices are available:

1. By email to the BODC contact for a particular project. For SIMORC they are Corallie Hunt (cora@bodc.ac.uk) and Lesley Rickards (ljr@bodc.ac.uk). Please note BODC currently has a limit of 5 MB for single email transfers.

2. By mail on DVD, CDROM, or diskette (Zip or floppy).

3. Data can be left on an accessible FTP site for BODC staff to collect. Please provide collection details to BODC.

4. By FTP to the BODC area of the Proudman Oceanographic Laboratory (POL) web site. There are data security issues relating to transfer by this method and ideally it should only be used as a last resort. A username and password is required. Suppliers wishing to provide data in this manner should contact BODC for details. **It is important** that users notify BODC **before** transfer of data in this way.

2.2 Incoming Data Formats

BODC can handle data in virtually any format, providing software to read it is readily available or that it is described in sufficient detail for us to write the software. In all cases, we require an explanation of how the format has been used so that we can understand what we have been given. Electronic submission may be eased by using a WinZip compatible compression routine. Statistical information, such as a list of file names supplied and their sizes or even the range of values for each parameter will help us ingest your data correctly.

Please pay particular attention to providing us with clear descriptions of the parameters that you have sent to us, including clear column headings and the units used. Indicate which parameters are directly measured and which are derived from a combination of measurements. For derived measurements, please include the formulae used by leaving them in a Microsoft Excel spreadsheet cell, including them in an accompanying document or providing a literature reference.

As we use UNIX systems we would appreciate it if filenames did not include embedded blanks. Please replace these with underscores (e.g. 'my_file' instead of 'my file').

2.3 Accompanying Metadata

BODC endeavour to incorporate all data submitted into relational database systems, 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 we require information about

- 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)

This information may be supplied in any standard document format (e.g. Microsoft Word or text) and will be incorporated into either specific metadata fields in our database or as comments in the documentation we will prepare to accompany your data.

3.0 Overview of BODC Data Processing Procedures

3.1 Summary

Moored instrument data go through several steps at BODC before they are incorporated in the National Oceanographic Database (NODB). Our aims are to ensure the data are of a consistent standard and to guarantee their long term security and utilisation. The data banking procedure involves reformatting of data files, quality control, entering information into Oracle tables, compiling documentation, and checking. All processes must be completed satisfactorily before the files can be archived in the database.

The NODB consists of a series of metadata inventories/tables which are held in an Oracle RDBMS (Relational Database Management System) and NetCDF data files which are held separately in a LINUX archive system. The data processing steps are outlined in the flow chart (Figure 1) and are explained more fully below.

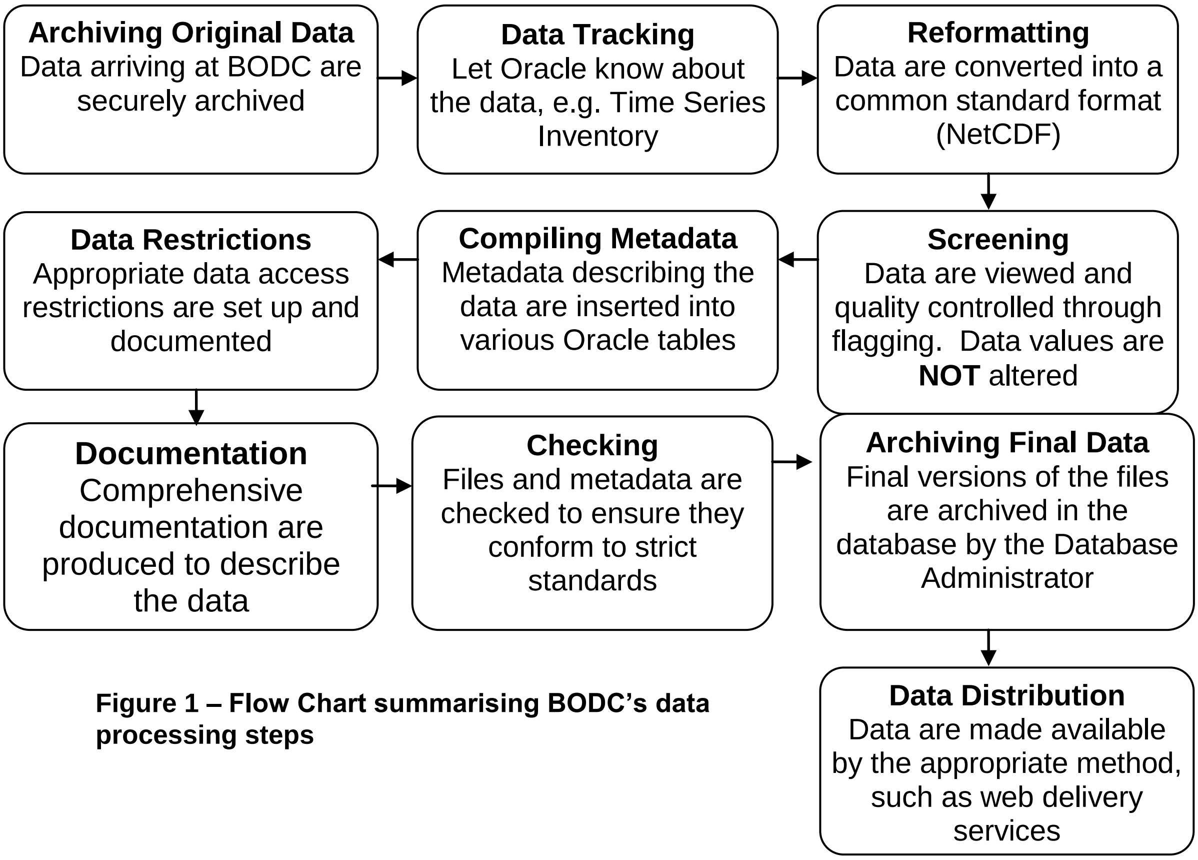


Figure 1 – Flow Chart summarising BODC's data processing steps

3.2 Archiving Original Data

When data are received at BODC they go through an Accession Procedure. The data files are archived securely in their original form along with any associated documentation. Information describing the data is added to Oracle tables so we can keep track of all data obtained.

BODC accepts data in most formats provided that the format is adequately described and that mandatory metadata are included. Most data are received in some kind of ASCII format, though some are received in binary formats such as MATLAB (.mat files). Data can be sent by various methods as mentioned in section 2.2.

3.3 Data Tracking

Data tracking procedures ensure that the Oracle database system know about the data. Moored instrument time series data that are known to BODC are catalogued in the Time Series Inventory. This includes one row per instrument and is not restricted to data held at BODC. It can also include instruments which were lost or failed and also data which is held by other organisations. It currently contains over 13400 entries from 86 organisations and is available for searching online.

3.4 Parameter Dictionary

The BODC Parameter Dictionary is used for labelling data as they are submitted to BODC. Instead of using non-standard descriptions for parameters, individual codes are assigned from the dictionary. The code gives 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 dictionary to about 9000 parameters. When BODC first started managing oceanographic data in the 1980s, we dealt with less than twenty parameters. This rapid increase in the number of parameters forced us to adopt a new approach to parameter management and develop the dictionary.

There are now dictionary entries for more than 18,000 physical, chemical, biological and geological parameters. Sometimes a single water bottle sample has been analysed for several hundred parameters. The dictionary is freely available and can be downloaded from:

www.bodc.ac.uk/data/codes_and_formats/parameter_codes/bodc_para_dict.html

Every Parameter is placed within a Group, linked by a 4-byte group code. To discover all Parameters within a particular Group, a search can be conducted using the group code. Groups are further classified into Categories to allow the user to focus a search starting at a very broad level. Examples of Categories are: acoustics, zooplankton, fatty acids. The top level of the hierarchy is the discipline (e.g. physics, chemistry, biology, etc.). Thus the hierarchy contains the following levels:

Discipline (at present 9 items)
Agreed Parameter Categories (at present 46 items)
BODC Parameter Groups (at present 341 items)
BODC Parameter Dictionary (at present 18855 items)

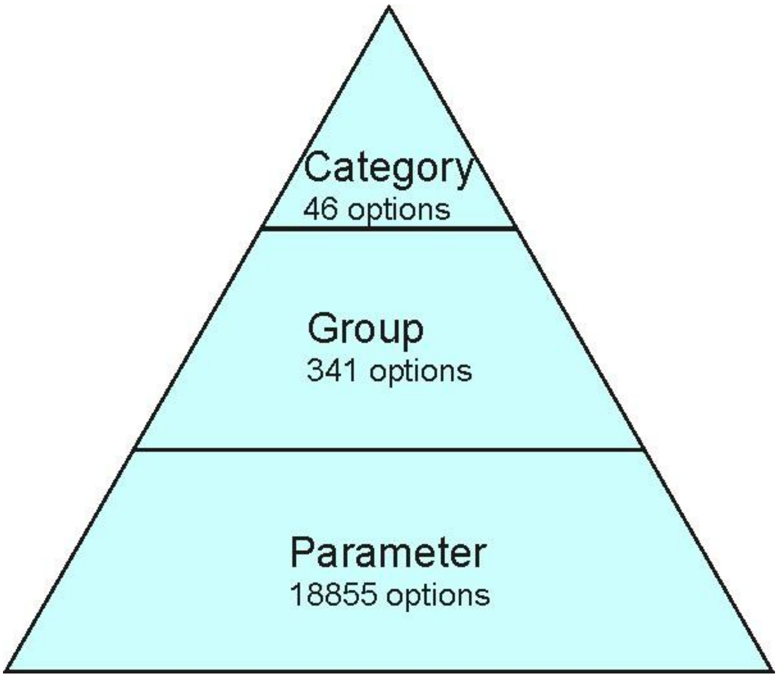


Figure 2: The data discovery hierarchy.

Each Category contains a subset of Groups. Each group contains a subset of Parameters. A user can focus a search for a Parameter or set of Parameters by navigating down the hierarchy.

3.5 Reformatting

As data arrive at BODC in various formats, the files need to be converted to a common standard format. It is important that all our data are held in one format as they can then be stored and distributed much more efficiently. It also ensures that parameter codes, flags, units, absent data values, etc. are consistent between files from different sources.

The format used for storing data at BODC is QXF (http://www.bodc.ac.uk/data/codes_and_formats/qxf_format/). QXF is based on the internationally recognised exchange format (NetCDF). It has the advantage of being able to handle multi-dimensional data from instruments such as moored Acoustic Doppler Current Profilers (ADCPs) and thermistor chains in one file. It is also platform independent and as it is array-based can be directly manipulated by MATLAB and other software packages. For further information on NetCDF, see http://my.unidata.ucar.edu/content/software/netcdf/index.html.

MATLAB software, including the NetCDF and Database Toolboxes, is used for the reformatting of data. Although the main code relating the transfer procedure is generic, a new module is written to read in the data for each new format received. Therefore, it is very helpful if data are sent in a standardised format where possible as this makes the transfer procedure much more efficient. Eight-byte parameter codes from the BODC Data Dictionary are assigned to each data channel, data are converted to standard units and absent data values are set. If any quality control flags have already been applied by the data originator, these are kept in the QXF file.

Metadata (e.g. positions, depths) from file headers or additional files are also extracted to be loaded into Oracle tables later.

The transfer process is also the first line of quality control defence. Often problems or inconsistencies with the data/metadata are picked up at this stage as many checks are built in to the transfer software. Some automated data screening may also occur at this stage, such as the flagging of out-of-range values.

3.6 Screening

All data are manually screened at BODC by a data scientist using in-house visualisation software (Edserplo). This software can be used to display moored data as a time series. Current meter data are also be plotted as scatter plots and profiles can be plotted for ADCP (Acoustic Doppler Current Profiler) data.

Quality control is carried out through flagging. Parameters can be plotted concurrently and records from different instruments can be compared. Data values are NOT changed or removed, but may be flagged if they appear suspect. Data values that the originator regards as suspect are flagged with 'L'. BODC uses 'M' to flag suspect data. 'N' flag is used to indicate absent data. 'P' is used to indicate calm conditions (for e.g. wave height data), 'Q' is used to indicate indeterminate; for example for wave period data which cannot be satisfactorily determined during calm conditions.

3.7 Compiling Metadata

The metadata (information about the data), such as collection date and times, mooring position, instrument type, instrument depth and sea floor depth are loaded into Oracle relational database tables. These fields of the main “series header” table for the National Oceanographic database are described in Annex 1. The values are carefully checked for errors and consistency with the data. The data originators may be contacted if any problems cannot be clearly resolved. Each data file has one row in the primary metadata table which is then linked to other tables that contain metadata relating more specifically to moorings, fixed stations, projects, cruises, etc.

3.8 Data Restrictions

Data may be restricted for a specified period or until a set date after the end of a project. It is important to ensure appropriate data access restrictions are set up to avoid data being supplied to a third party during the restricted period. Exceptions may be made for scientists working on the project or if authorisation has been given from the data originator to supply the data. All SIMORC data held in the BODC database is set as restricted as we do not supply this data directly. The data may be obtained from the official SIMORC web site.

3.9 Documentation

Comprehensive documentation is compiled to accompany the data. All data sets need to be fully documented to ensure they can be used in the future without any ambiguity or uncertainty. The documentation is compiled using information supplied by the data originator (e.g. data reports, comments on data quality) and any further information gained by BODC during screening. It will include information such as mooring details, instrumentation, data quality, calibration and processing carried out by the data originator, and BODC processing and quality control. If any particular major problems are associated with the data, these can be described in a Problem Report. General comments are stored in Data Quality Reports.

A document management system is used to load and link the documents which are stored as tagged XHTML in Oracle tables.

A Series Metadata Report for each data file containing all of its linked documentation is available on the internet at http://www.bodc.ac.uk/data/documents/series/ishref/, using Web Services, where ishref is the data series reference number.

e.g. http://www.bodc.ac.uk/data/documents/series/702665/

3.10 Quality Checking

Before the data set can be loaded to the database, the QXF files and Oracle metadata are thoroughly checked to ensure they conform to stringent BODC standards. MATLAB software is used through a GUI interface to do this. Checks include:

- Channel limits are within the prescribed range for the parameter or are flagged if they are out with the range
- Time channel is increasing
- Null flags correspond with absent data values
- File doesn't begin or end with null data cycles
- Start and end time in data file and Oracle table agree
- No problems with QXF data file format
- Data series linkages
- Checks for gaps in data

Google Earth is also used to check positions. The screening and documentation are then also audited by a second person, as this can be helpful in highlighting any inconsistencies in flagging.

3.11 Archiving Final Data

Once the files have been screened and checked, the archiving of the final data set is performed by the BODC Database Administrator. The files are renamed and archived into the file system and the Oracle tables are updated to indicate the banked status of the data.

3.12 Data Distribution and Delivery

The data can then be distributed by a method appropriate for the project, such as through web delivery services or on CD ROM. If the user requires data in an alternative format, the data can be translated into a specific ASCII formats or other NetCDF formats, such as WOCE NetCDF before supplying.

In the case of SIMORC, the data will not be available from BODC, as it will be made available by MARIS via the SIMORC website. Any enquirers requesting the data from BODC will be redirected to the website.

4.0 Current Meter Data - Quality Control

4.1 Checklist of Metadata Required for Processing/QC/Documentation

The checklist and example information below shows the information used by BODC to ensure that the data are adequately described.

Owner Details	
Name of country responsible for data	e.g. UK
Name of organisation responsible for data	e.g. POL
Project Name (if applicable)	e.g. Coastal Observatory
Data Type (e.g. current, wave, sea level, met)	e.g. Current Meter
Mooring/Instrument Details	
Instrument category (e.g. current meter, wave recorder)	e.g. Paddle wheel current meter
Mooring/Rig Number	e.g. 1234
Instrument model and manufacturer	e.g. Aanderaa RCM7
Principle of measurement?	e.g. Vector averaged currents
Any instrument modifications?	e.g. N/A
Additional notes on mooring structure (e.g. buoyancy measures, number of sensors)	e.g. In total 4 current meters on the mooring, subsurface buoyancy
Additional notes on performance of mooring (incl. condition on recovery, events which may have affected the data etc.)	e.g. weed found close to current meter, possible fouling of data
Latitude of mooring (degrees)	e.g. 53.85°
Longitude of mooring (degrees)	e.g. -3.26°
Time zone	e.g. GMT/UTC
Site Area and Name of Site	e.g. Irish Sea, ABC site
Method of position fix	e.g. GPS
Water column Depth (m)	e.g. 352m
Sea Floor Depth Qualifier (e.g. echo sounder, GEBCO)	e.g. GEBCO
Depth of meter or shallowest sensor (e.g. ADCP bin)	e.g. 300m
Depth of deepest sensor (can be same as above) (m) – (Specific for ADCPs and thermistors)	e.g. N/A
Interval depth between successive bins (m) – (Specific for ADCPs and thermistors)	e.g. 20m
Series Depth Field Qualifier? (e.g. Sea floor reference etc)	e.g. Mean Sea Level
Timing Details	
Date and Time of Deployment (UTC)	e.g. 25/08/04, 10:00
Date and Time of start of usable data (UTC)	e.g. 25/08/04, 10:20
Date and Time of Recovery (UTC)	e.g. 14/04/05, 16:30
Date and Time of end of usable data (UTC)	e.g. 14/04/05, 16:10
Nominal time interval between successive data cycles in series (seconds)	e.g. 600s

Type of sampling (e.g. instantaneous, averaged)	Current direction – vector averaged Temperature – instantaneous
Parameter Details	
Parameters measured and any definitions where the parameter is not obvious (See 3.3. BODC Parameter Dictionary below) Data Processing Details	e.g. Current speed (LCSAEL01), current direction (LCDAEL01), temperature (TEMPPR01), salinity (PSALPS01), pressure (PREXPR01)
Originator's Data Format Description of calibrations Description of any data processing that has occurred (manufacturers and in-house)	e.g. ASCII, .mat

4.2 BODC Parameter Dictionary codes

To get a comprehensive list of our parameter codes and their definitions, you can go to our online parameter dictionary at:
http://www.bodc.ac.uk/data/codes_and_formats/parameter_codes/bodc_para_dict.html

From here you can download the Parameter code list, the Parameter Group code list and the Units of Measurement code list, among others.

To aid your search we have included the most frequently used Parameter Group codes and Parameter codes for current meter data below:

Frequently Used Parameter Group Codes for Current Meter Data:

Parameter Group Code	Description
RFVL	Parameters expressing the velocity (including scalar speeds and directions) of water column horizontal movement, commonly termed currents
LRZA	Vertical water movement parameters including non-directional speeds
TEMP	Temperature of the water column
PSAL	Salinity of the water column

Frequently Used Parameter code for Current Meter Data:

Parameter Code	Description
LCNSZZ01	Northward current velocity in the water column
LCNSEL01	Northward current velocity (Eulerian) in the water column by in-situ current meter
LCNSLA01	Northward current velocity (Lagrangian) in the water column by tracked drifting buoy
LCNSAP01	Northward current velocity (Eulerian) in the water column by moored acoustic doppler current meter
LCNSAS01	Northward current velocity (Eulerian) in the water column by

	shipborne acoustic doppler current profiler
LCEWZZ01	Eastward current velocity in the water column
LCEWEL01	Eastward current velocity (Eulerian) in the water column by in-situ current meter
LCEWLA01	Eastward current velocity (Lagrangian) in the water column by tracked drifting buoy
LCEWAP01	Eastward current velocity (Eulerian) in the water column by moored acoustic doppler current meter
LCEWAS01	Eastward current velocity (Eulerian) in the water column by shipborne acoustic doppler current profiler
LCDAZZ01	Current direction in the water column
LCDAEL01	Current direction (Eulerian) in the water column by in-situ current meter and correction to true North
LCDAEL02	Current direction (Eulerian) in the water column by in-situ current meter and correction to true North
LCDALA01	Current direction (Lagrangian) in the water column by tracked drifting buoy
LCDAAP01	Current direction (Eulerian) in the water column by moored acoustic doppler current meter and correction to true North
LCDAAP02	Current direction (Eulerian) in the water column by moored acoustic doppler current meter and correction to true North
LCDAAS01	Current direction (Eulerian) in the water column by shipborne acoustic doppler current profiler and correction to true North
LCSAZZ01	Current speed in the water column
LCSAEL01	Current speed (Eulerian) in the water column by in-situ current meter
LCSAEL02	Current speed (Eulerian) in the water column by in-situ current meter
LCSALA01	Current speed (Lagrangian) in the water column by tracked drifting buoy
LCSAAP01	Current speed (Eulerian) in the water column by moored acoustic doppler current meter
LCSAAP02	Current speed (Eulerian) in the water column by moored acoustic doppler current meter
LCSAAS01	Current speed (Eulerian) in the water column by shipborne acoustic doppler current profiler
RFVLDR01	Downstream current velocity in a river by direct reading current meter

4.2.1 Web Services

BODC's parameter vocabulary can be accessed using web services.

A web service is a collection of protocols and standards used for exchanging data between applications or systems. Software applications written in various programming languages and running on various platforms can use Web services to exchange data over the Internet, in a manner similar to inter-process communication on a single computer. This interoperability (e.g. between Java and Python, or Windows and Linux applications) is due to the use of open standards.

Further information is available from the BODC Web Services home page: http://www.bodc.ac.uk/products/web_services/

4.3 Glossary

- **Acoustic Doppler Current Profiler (ADCP)** – A current meter which measures current direction in 3 dimensions. As opposed to the average current meter, an ADCP can measure current speeds and direction at varying depths using a principle known as the Doppler Shift
- **Bin (ADCP) –** A depth where an ADCP will measure current direction and speed using the Doppler Shift principle. An ADCP will measure current speeds and directions at a number of 'bins' above or below it, which are placed at regular depth intervals. Data gathered at those 'bins' furthest away from the instrument is generally of a poorer quality than that data collected at closer 'bins'
- **Channel** – Generally represents measurements of a particular parameter (e.g. temperature): the term channel is used in preference to parameter because there can be more than one measurement of a particular parameter within a datacycle
- **Current** – Regular flow of a mass of water. The flow is generally consistent and is dependent on the properties of the water mass
- **Mooring Knockdown** – High speed currents can cause the mooring to 'fall over'. This can be seen in the time series as an increase in pressure, a result of the sensor being pushed down to a deeper depth than it was positioned at
- **Mooring/Rig** – deployment holding various oceanographic instruments including current meters, ADCPs, thermistors, wave riders etc
- **Trawling** – This is a result of a ship/boat catching the subsurface buoyancy of the mooring carrying it from its mooring location
- **u, v components** – breakdown of velocity – northerly travel and easterly travel
- **Velocity** – 2 dimensional property of a moving object, incorporating the speed at which it travels and the direction in which it is travelling

4.4 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, 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.

4.4.1 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 of detecting 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 3 – Time Series plot showing a 'spike' in salinity and its appropriate flag

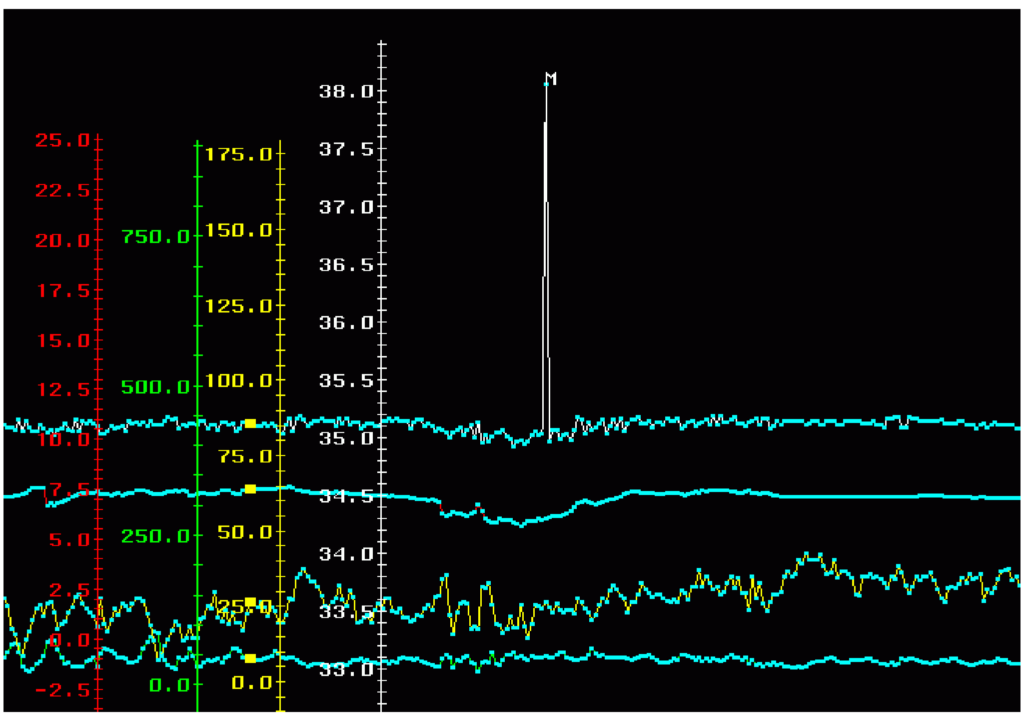


Figure 3 gives an example of a 'spike' in salinity within a series. The spike has been flagged as suspect with the customary 'M' flag. As can be seen only one point is out-of-range indicating it is likely to be an instrumental error.

Figure 4 – Time series plot showing 'noisy' temperature, salinity and speed data

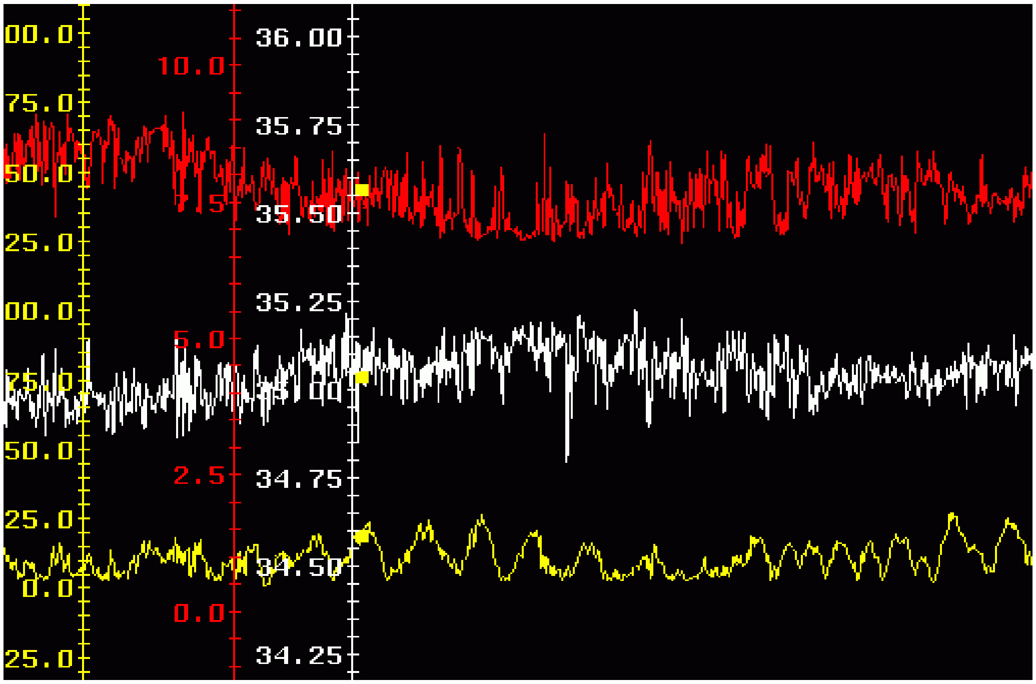
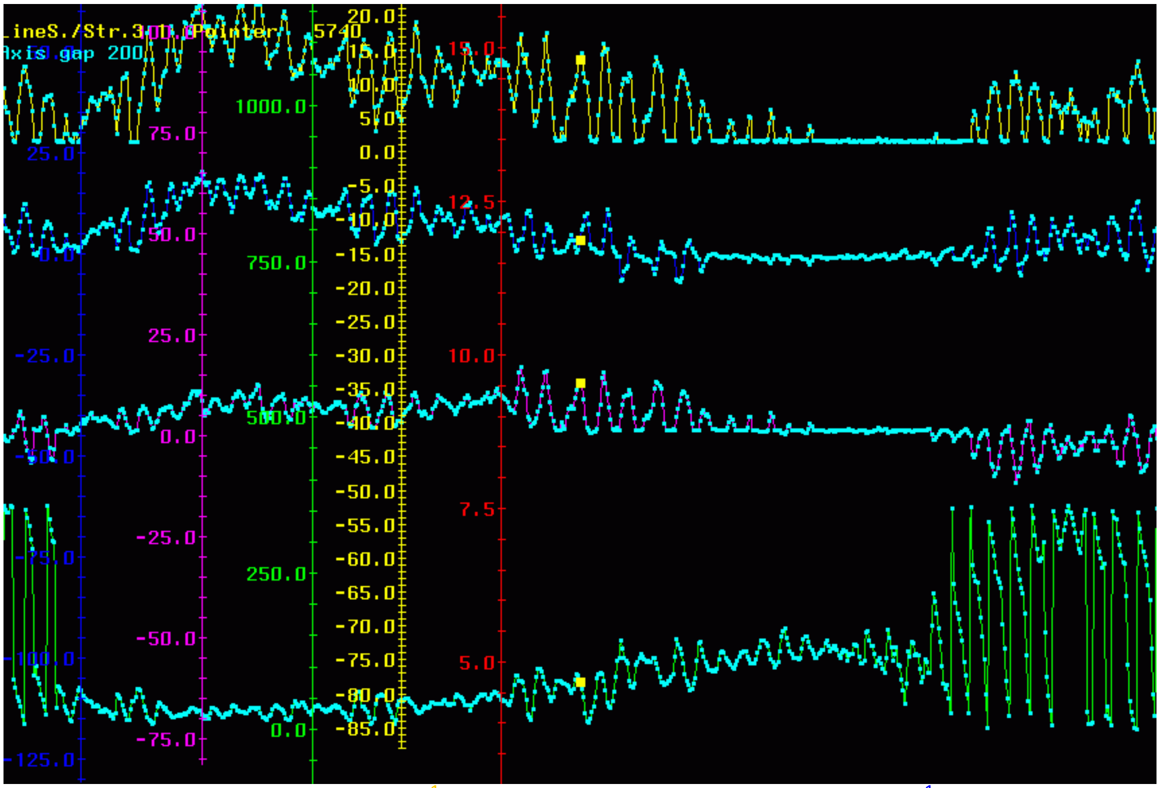


Figure 4 is an example of a situation where it is difficult to identify spikes because the surrounding data are noisy indicating a natural phenomenon having an effect on the localised water conditions.

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 5 - 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 5 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 3).

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 checked 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).

4.4.2 Data Limit Tests

Below are a range of tests taken from the IOC/CEC's 'Manual of Quality Control Procedures for Validation of Oceanographic Data', 1993, which we can use to gauge whether the data is realistic or not:

a) Gross Error Limits

i) Current Speed
Current speeds should not exceed the maximum speed which the current meter can measure based on the sampling period and scaling factor used, or $4ms^{-1}$, whichever is the smaller. The current speed should not be negative.

ii) Current Direction
All current directions should lie between 000º and 360º

iii) Temperature
All temperatures should lie within the range of the sensor

iv) Conductivity/Salinity
All conductivity values should lie within the range of the sensor

b) Rate of Range Checks

i) Current speed and direction
Rate-of-change checks for current speed and direction are best applied to orthogonal components of the current velocity, since these can be considered to be cosine functions with definable expected differences between sampling points.
The theoretical differences between two consecutive current speed samples u_1 and u_2 for various sampling intervals (Δt), assuming a smooth sinusoidal semi-diurnal tidal current with a period of 12.42 hours, are given below:

Δt (min)	Theoretical $|u_1 - u_2|$	Factor	Allowable $|u_1 - u_2|$
5	**0.0422 u**	**2.0**	**0.08ms^{-1}**
10	**0.0843 u**	**1.8**	**0.15ms^{-1}**
15	**0.1264 u**	**1.6**	**0.20ms^{-1}**
20	**0.1685 u**	**1.5**	**0.25ms^{-1}**
30	**0.2523 u**	**1.4**	**0.35ms^{-1}**
60	**0.5001 u**	**1.2**	**0.60ms^{-1}**

:where *u* is the orthogonal tidal current amplitude. In order to allow for some inherent variability in current speed and direction signal and for asymmetric tidal current speed curves, these differences will have been increased by the above factors whilst *u* has been set at 1.0ms^{-1} since the variability will increase with decreasing *u*. The resulting allowable maximum difference between samples for particular sampling intervals is provided above.

ii) Sea Temperature

$$\frac{|T_1 - T_2| \leq \Delta t}{60}$$

:where T_1 and T_2 are consecutive temperature measurements and Δt is the sampling interval in minutes.

iii) Conductivity/Salinity

$$\frac{|S_1 - S_2| \leq \Delta t}{60}$$

:where S_1 and S_2 are consecutive salinity measurements and Δt is the sampling interval in minutes.

c) Stationarity Checks

The occurrence of constant values of data depends on the variable being measured, the sampling interval used and the resolution of the sensor.

i) Current Speed

Constant current speeds are uncommon although, theoretically, two consecutive values may be the same. A flag is set against each current speed which is equal in value to the two previous values, regardless of the sampling interval.

ii) Current Direction

Almost constant directions may be generated by topographic effects. The following numbers of consecutive equal values are allowed depending on sampling interval:

Δt (min)	**Number of Consecutive Equal Values**
5	**12**
10	**6**
15	**4**
20	**3**
30	**2**
60	**2**

A flag should be set against each current direction point which is equal in value to the previous 12, 6, 4, 3 or 2 previous values (as applicable).

iii) Temperature

Constant temperature values are relatively common in well-mixed water, and the number of consecutive equal values allowed is thus large, being:

$$24 \times \frac{60}{\Delta t\ (\text{min})}$$ (i.e. up to one day is allowed)

:where Δt is the sampling interval in minutes. A flag should be set against all data points that are preceded by at least a day of constant values.

iv) Conductivity/Salinity

Constant salinity values are also relatively common in well-mixed water and a similar stationarity check to that for temperature is applied:

$$24 \times \frac{60}{\Delta t\ (\text{min})}$$ (i.e. up to one day is allowed)

: where Δt is the sampling interval in minutes. A flag should be set against all data points which are preceded by at least a day of constant values.

4.4.3 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 flow-field 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 show 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 6 and 7:

Figure 6 – Example of satisfactory directional data from a data series

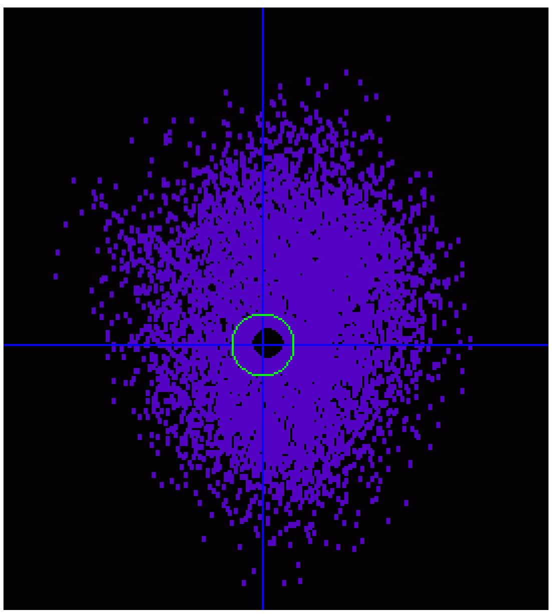
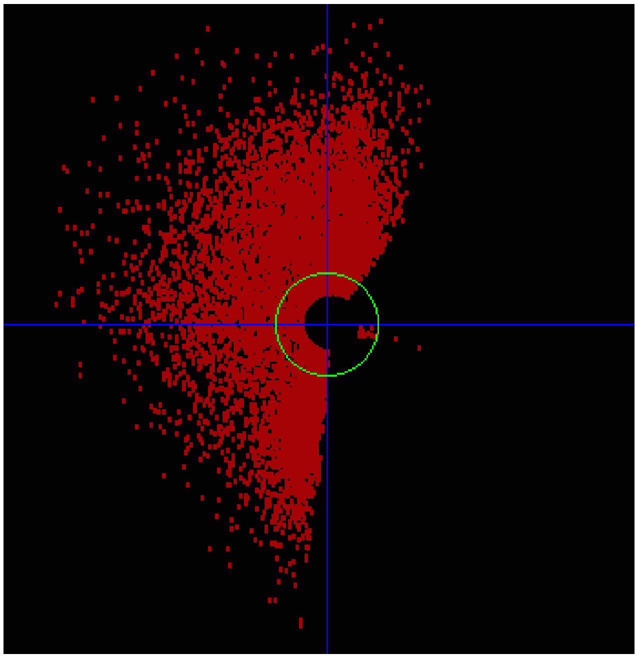


Figure 6 shows an example of a 'good' scatter plot. There seems to be an even array of current directions indicating that the compass was not being hindered.

However, Figure 7 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 7 – Example of unsatisfactory directional data from a data series



4.4.4 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 picked up easily. 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 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. These cause spikes in all parameters and are 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.

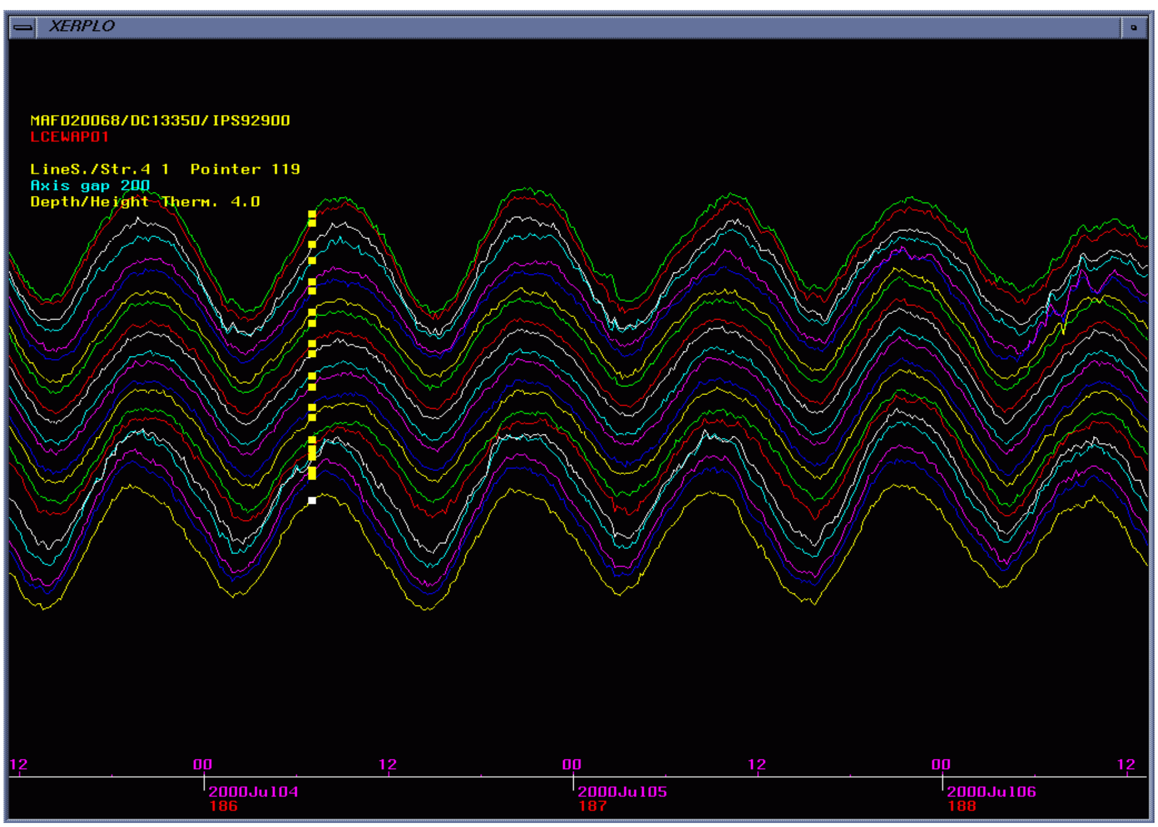
- Poor quality recording tape. This is indicated by the appearance of suspect data at regular intervals in all parameters.

- Sensor drift. This is a slow change in the response of the sensor.

4.4.5 Differences in Screening Procedure for ADCPs and Thermistors

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 8 below:

Figure 8 - ADCP profile showing the East-West component of the current direction



Thermistors chains are similar, but sample temperature at different bin levels. The screening procedure for these three dimensional data types is similar in principle to 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.

4.6 Accompanying Documentation

At BODC we include a set of standard documentation with every data series. Using in-house 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 are as follows:

- Data Activity – this document describes the 'event'. In the case of current meters it generally gives the dates of deployment, description of the location of the mooring and its setup, including the latitude and longitude, and the depths and meter numbers of all the other sensors present on the mooring. 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.

- 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 over time.

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

- BODC Screening Document – This is a generic document linked to all datasets giving a brief summary of the screening procedure BODC undertakes 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.

5.0 Wave Data - Quality Control

5.1 Checklist of Metadata Required for Processing/QC/Documentation

The checklist and example information below shows the information used by BODC to ensure that the data are adequately described.

Owner Details	
Name of country responsible for data	e.g. UK
Name of organisation responsible for data	e.g. POL
Project Name (if applicable)	e.g. Coastal Observatory
Data Type (e.g. current, wave, sea level, met)	e.g. Wave Statistics
Mooring/Instrument Details	
Instrument category (e.g. current meter, wave recorder)	e.g. Wave recorder
Mooring/Rig Number	e.g. 1234
Instrument model and manufacturer	e.g. Datawell Waverider
Principle of measurement?	e.g. Accelerometer
Any instrument modifications?	e.g. N/A
Additional notes deployment	e.g. Any known data gaps / dropout problems
Additional notes on performance of mooring	e.g. high-frequency cut-off
Latitude of mooring (degrees)	e.g. 53.85°
Longitude of mooring (degrees)	e.g. -3.26°
Time zone	e.g. GMT/UTC
Site Area and Name of Site	e.g. Irish Sea, ABC site
Method of position fix	e.g. GPS
Water column Depth (m)	e.g. 352m
Sea Floor Depth Qualifier (e.g. echo sounder, GEBCO)	e.g. GEBCO
Depth of meter or shallowest sensor (e.g. ADCP bin)	e.g. 300m
Timing Details	
Date and Time of Deployment (UTC)	e.g. 25/08/04, 10:00
Date and Time of start of usable data (UTC)	e.g. 25/08/04, 10:20
Date and Time of Recovery (UTC)	e.g. 14/04/05, 16:30
Date and Time of end of usable data (UTC)	e.g. 14/04/05, 16:10
Nominal time interval between successive data cycles in series (seconds)	e.g. 600s
Type of sampling (e.g. instantaneous, averaged)	
Parameter Details	
Parameters measured and any definitions where the parameter is not obvious (See 5.2. BODC Parameter Dictionary below)	e.g. significant wave height (GTDH%), zero upcrossing period (GTZA%)
Data Processing Details	

Originator's Data Format Description of calibrations Description of any data processing that has occurred (manufacturers and in-house)	e.g. ASCII, .mat

5.2 BODC Parameter Dictionary codes

To get a comprehensive list of our parameter codes and their definitions, you can go to our online parameter dictionary at:
http://www.bodc.ac.uk/data/codes_and_formats/parameter_codes/bodc_para_dict.html

From here you can download the Parameter code list, the Parameter Group code list and the Units of Measurement code list, among others.

To aid your search we have included the most frequently used Parameter Group codes and Parameter codes for current meter data below:

Frequently Used Parameter Group Codes for Wave Data:

Parameter Group Code	Description
GTDH	Parameters expressing the significant wave height
GTZA	Zero upcrossing period
GCMX	Maximum recorded wave height
GTPK	Spectral peak wave period.

Frequently Used Parameter Codes for Wave Data:

Code	Description
GPEDFA01	Spectrum peak energy direction of waves on the water column by waverider and Fourier analysis
GPSPFA01	Directional spreading at spectral peak of waves on the water column by waverider and Fourier analysis
GSPKFA01	Spectral peakedness factor of waves on the water column by waverider and Fourier analysis after Goda (1970)
GSWDFA01	Spectral width of waves on the water column by waverider and computation from moments as defined by Cartwright and Longuet-Higgins (1956)
GTPKFA01	Spectral maximum period of waves on the water column by waverider and Fourier analysis
GAVGEV01	Average height of waves on the water column by visual estimation
GAVGVA01	Average height of waves on the water column by waverider
GAVHCA01	Average height of waves (highest 1/3rd) on the water column by waverider and automated record analysis
GAVHCS01	Average height of waves (highest 1/3rd) on the water column by wavestaff and automated record analysis

GAVHMA01	Average height of waves (highest 1/3rd) on the water column by waverider and manual record analysis
GCARFA01	Characteristic height of waves {Hrms*4} on the water column by waverider and Fourier analysis
GCARMA01	Characteristic height of waves {Hrms*4} on the water column by waverider and manual record analysis
GCMXCA01	Maximum height of waves on the water column by waverider and automated record analysis
GCMXVA01	Maximum height of waves on the water column by waverider
GHSWEV01	Height of swell waves on the water column by visual estimation
GMNLCA01	Minimum level of waves on the water column by waverider and automated record analysis
GMX1CA01	Maximum height of waves (in 1-hour period) on the water column by waverider and automated record analysis
GMXACA01	Maximum level plus minimum level of waves on the water column by waverider and automated record analysis
GMXHCA01	Maximum height of waves (in 3-hour period) on the water column by waverider and automated record analysis
GMXHFA01	Maximum height of waves (in 3-hour period) on the water column by waverider and Fourier analysis
GMXHMA01	Maximum height of waves (in 3-hour period) on the water column by waverider and manual record analysis
GMXHTD01	Maximum height of waves (in 3-hour period) on the water column by water level recorder pressure sensor
GMXLCA01	Maximum level of waves on the water column by waverider and automated record analysis
GPSWEV01	Swell period of waves on the water column by visual estimation
GRMSCA01	Root mean square displacement of waves on the water column by waverider and automated record analysis
GRMSVA01	Root mean square displacement of waves on the water column by waverider
GTCACA01	Average crest period of waves on the water column by waverider and automated record analysis
GTCACP01	Average crest period of waves on the water column by water level recorder pressure sensor and automated record analysis
GTCAEV01	Average crest period of waves on the water column by visual estimation
GTCAZZ01	Average crest period of waves on the water column
GTDHCA01	Significant height of waves {Hs} on the water column by waverider and automated record analysis
GTDHCS01	Significant height of waves {Hs} on the water column by wavestaff and automated record analysis
GTDHFA01	Significant height of waves {Hs} on the water column by waverider and Fourier analysis
GTDHFP01	Significant height of waves {Hs} on the water column by water level recorder pressure sensor and Fourier analysis
GTDHTD01	Significant height of waves {Hs} on the water column by pressure sensor
GTDHUP01	Significant height of waves {Hs} on the water column by water level recorder pressure sensor and Fourier analysis with NO depth

	attenuation correction
GTDHVA01	Significant height of waves {Hs} on the water column by waverider
GTDHZZ01	Significant height of waves {Hs} on the water column
GTKCMA01	Second maximum level of waves on the water column by waverider and manual record analysis
GTKCMP01	Second maximum level of waves on the water column by water level recorder pressure sensor and manual record analysis
GTKDMA01	Second minimum level of waves on the water column by waverider and manual record analysis
GTKDMP01	Second minimum level of waves on the water column by water level recorder pressure sensor and manual record analysis
GTZACA01	Average zero crossing period of waves {Tz} on the water column by waverider and automated record analysis
GTZACP01	Average zero crossing period of waves {Tz} on the water column by water level recorder pressure sensor and automated record analysis
GTZACS01	Average zero crossing period of waves {Tz} on the water column by wavestaff and automated record analysis
GTZAFA01	Average zero crossing period of waves {Tz} on the water column by waverider and Fourier analysis
GTZAFB01	Average zero crossing period of waves {Tz} on the water column by shipborne wave recorder and Fourier analysis
GTZAFP01	Average zero crossing period of waves {Tz} on the water column by water level recorder pressure sensor and Fourier analysis
GTZATD01	Average zero crossing period of waves {Tz} on the water column by pressure sensor
GTZAUP01	Average zero crossing period of waves {Tz} on the water column by water level recorder pressure sensor and Fourier analysis with NO depth attenuation correction
GTZAVA01	Average zero crossing period of waves {Tz} on the water column by waverider
GTZAZZ01	Average zero crossing period of waves {Tz} on the water column
GTZHCA01	Average zero crossing period (highest one third) of waves on the water column by waverider and automated record analysis
GTZHCS01	Average zero crossing period (highest one third) of waves on the water column by wavestaff and automated record analysis
GTZMCA01	Maximum zero crossing period of waves on the water column by waverider and automated record analysis
GTZMCS01	Maximum zero crossing period of waves on the water column by wavestaff and automated record analysis
GZMXCA01	Maximum zero crossing height of waves on the water column by waverider and automated record analysis
GZMXCS01	Maximum zero crossing height of waves on the water column by wavestaff and automated record analysis
GZMXFA01	Maximum zero crossing height of waves on the water column by waverider and Fourier analysis
GDSWEV01	Direction of swell waves on the water column by visual estimation
GDSWZZ01	Direction of swell waves on the water column
GSPRFA01	Directional spread of waves on the water column by waverider and computation from cross spectra
GWDREV01	Direction of waves on the water column by visual estimation

GWDRZZ01	Direction of waves on the water column

5.2.1 Web Services

BODC's parameter vocabulary can be accessed using web services.

A web service is a collection of protocols and standards used for exchanging data between applications or systems. Software applications written in various programming languages and running on various platforms can use Web services to exchange data over the Internet, in a manner similar to inter-process communication on a single computer. This interoperability (e.g. between Java and Python, or Windows and Linux applications) is due to the use of open standards.

Further information is available from the BODC Web Services home page: http://www.bodc.ac.uk/products/web_services/

5.3 Glossary

- **Hs : Significant wave height** - measure of the average height of the highest one third of waves during the record in metres
- **Tz: Zero upcrossing period** - the mean wave period (taken as the average time between consecutive crossings of the mean sea level line in an upwards direction) in seconds
- **Tmax: Maximum wave height** - maximum recorded wave height in metres
- **Tcrest: Crest period** - time taken between consecutive crossings of wave crests
- **Tpeak:** Peak period also known as dominant wave period - it is the period corresponding to the frequency band with the maximum value of spectral density

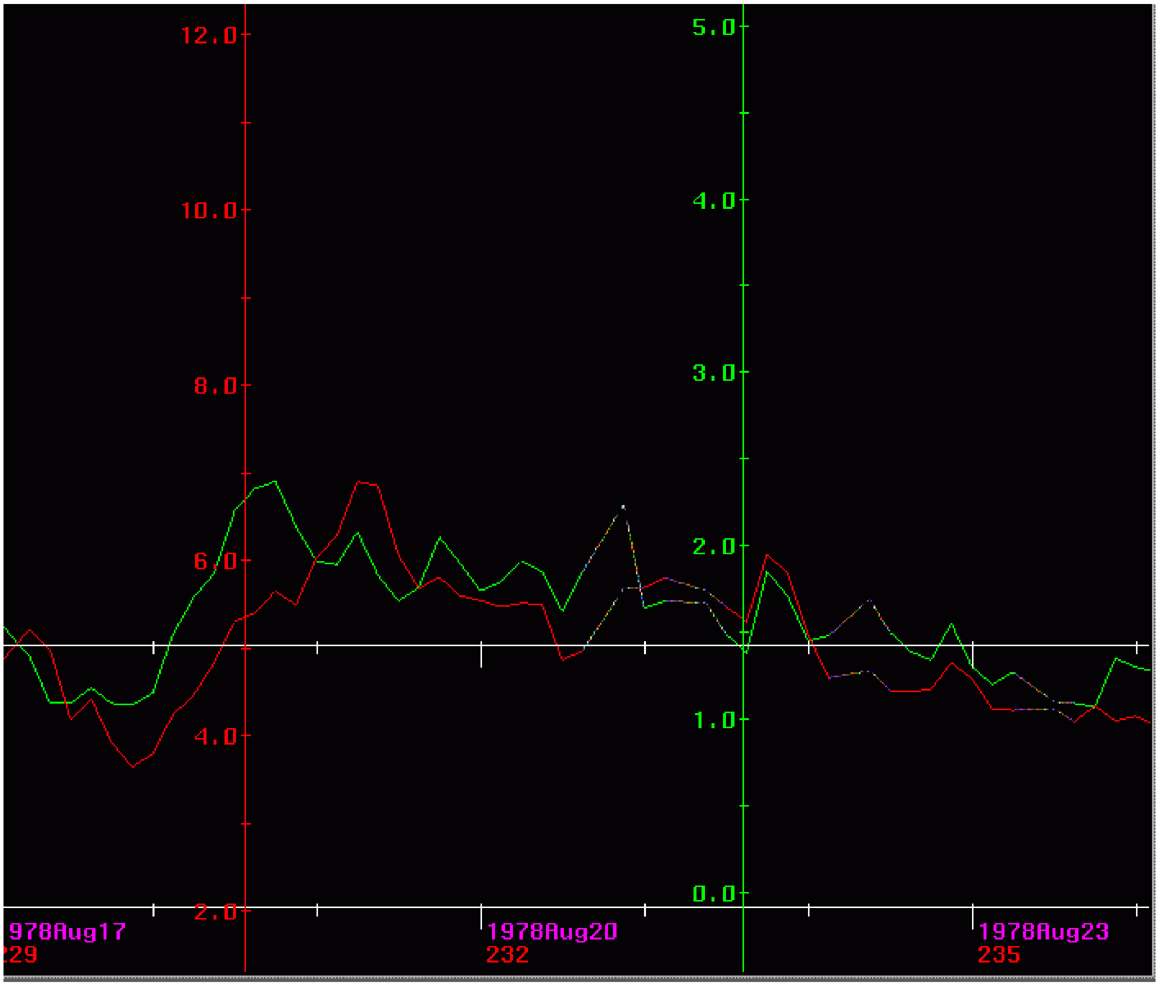
5.4 Screening and QC Procedures – All Wave Data

After reformatting using BODC's in house Transfer system, data will be screened on a graphics workstation. BODC's in-house software for quality controlling current meter data comprises a visualisation tool called SERPLO, developed in response to the needs of BODC whose requirement involved the rapid inspection and non-destructive editing of large volumes of data. SERPLO allows the user to select specific data sets and to view them in various forms, to visually asses their quality. 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 might not otherwise be picked up if just viewing large columns of figures. 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. These flags do not change the data; they purely highlight potential problems with the data, allowing the end user to make the decision as to whether to use the data or not. 'P' is used to indicate calm conditions (for e.g. wave height data), 'Q' is used to indicate indeterminate; for example for wave period data which cannot be satisfactorily determined during calm conditions

5.4.1 Time Series Plots

Using the time series plot 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 9 – 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:

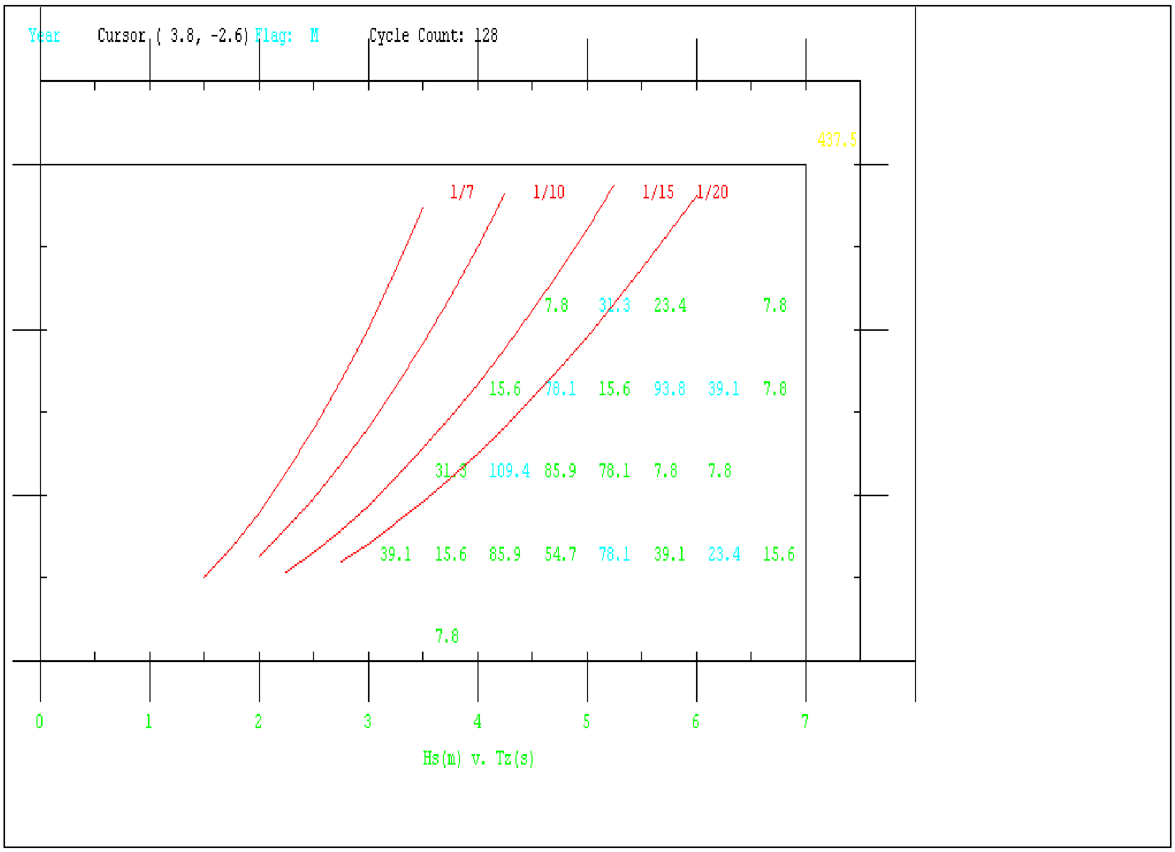
- Instrument failure test (10 or more consecutive points of identical value)
- Wandering mean test (interval between successive upcrossings of >25 seconds)
- Check that Tz falls within the range 2-16 seconds
- Check that Tpeak falls within the range 3 -20 seconds
- Check that Tz not less than Tcrest
- 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 ie 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

5.4.2 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 10 – 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 (NB 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

5.4.3 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.

5.4.4 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

5.5 Problems Associated with Wave Data

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

5.6 Accompanying Documentation

At BODC we include a set of standard documentation with every data series. Using in-house 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 are as follows:

- Data Activity – this document describes the 'event'. In the case of wave buoys or similar it generally gives the dates of deployment and 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 that 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.

- 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 that 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.

- BODC Screening Document – This is a generic document linked to all datasets giving a brief summary of the screening procedure BODC undertakes 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.

6.0 Sea Level Data – Quality Control

6.1 Checklist of Metadata Required for Processing/QC/Documentation

The checklist and example information below shows the information used by BODC to ensure that the data are adequately described.

Owner Details	
Name of country responsible for data	e.g. UK
Name of organisation responsible for data	e.g. POL
Project Name (if applicable)	e.g. Coastal Observatory
Data Type (e.g. current, wave, sea level, met)	e.g. pressure/water level
Mooring/Instrument Details	
Instrument category (e.g. current meter, wave recorder)	e.g. Water Level Recorder
Mooring/Rig Number	e.g. 1234
Instrument model and manufacturer	e.g. Aanderaa WLR
Principle of measurement?	e.g. pressure sensor
Any instrument modifications?	e.g. N/A
Additional notes deployment	e.g. Any known data gaps / dropout problems
Additional notes on performance of mooring	e.g. high-frequency cut-off
Latitude of mooring (degrees)	e.g. 53.85°
Longitude of mooring (degrees)	e.g. -3.26°
Time zone	e.g. GMT/UTC
Site Area and Name of Site	e.g. Irish Sea, ABC site
Method of position fix	e.g. GPS
Water column Depth (m)	e.g. 352m
Sea Floor Depth Qualifier (e.g. echo sounder, GEBCO)	e.g. GEBCO
Depth of meter or shallowest sensor (e.g. ADCP bin)	e.g. 300m
Depth of deepest sensor (can be same as above) (m) – (Specific for ADCPs and thermistors)	e.g. N/A
Interval depth between successive bins (m) – (Specific for ADCPs and thermistors)	e.g. 20m
Series Depth Field Qualifier? (e.g. Sea floor reference etc)	e.g. Sea Floor Reference
Timing Details	
Date and Time of Deployment (UTC)	e.g. 25/08/04, 10:00
Date and Time of start of usable data (UTC)	e.g. 25/08/04, 10:20
Date and Time of Recovery (UTC)	e.g. 14/04/05, 16:30
Date and Time of end of usable data (UTC)	e.g. 14/04/05, 16:10
Nominal time interval between successive data cycles in series (seconds)	e.g. 600s

Type of sampling (e.g. instantaneous, averaged)	
Parameter Details	
Parameters measured and any definitions where the parameter is not obvious (See 6.2. BODC Parameter Dictionary below) Data Processing Details	e.g, pressure (PREXPR01)
Originator's Data Format Description of calibrations Description of any data processing that has occurred (manufacturers and in-house)	e.g. ASCII, .mat

6.2 BODC Parameter Dictionary codes

To get a comprehensive list of our parameter codes and their definitions, you can go to our online parameter dictionary at:
http://www.bodc.ac.uk/data/codes_and_formats/parameter_codes/bodc_para_dict.html

From here you can download the Parameter code list, the Parameter Group code list and the Units of Measurement code list, among others.

To aid your search we have included the most frequently used Parameter Group codes and Parameter codes for current meter data below:

Frequently Used Parameter Group Codes for Sea Level Data:

Parameter Group Code	Description
AHGT	All vertical spatial parameters including depth, height and pressure when used as an independent variable.
ASLV	Measurements and predictions of the displacement of the water column surface from a fixed, stable reference
CAPH	Measurements of air pressure as the dependent variable (ie excluding air pressure measured to specify the z co-ordinate of a balloon or sonde), including derived parameters such as tendency (rate of change) and related parameters such as air density.
PREX	Measurements of the displacement of the water column surface from a fixed, stable reference expressed as either total pressure or sea pressure

Frequently Used Parameter code for Sea Level Data:

Parameter Code	Description
ADEPPS01	Depth below surface of the water column by semi-fixed in-situ pressure sensor and conversion to depth using unspecified algorithm
ASLVMSPG	Surface elevation (Mean Sea Level datum) of the water column by fixed in-situ pressure sensor
ASLVTD01	Surface elevation (unspecified datum) of the water column by fixed in-situ pressure sensor
ASLVZZ01	Surface elevation (unspecified datum) of the water column
PRESPS01	Pressure (measured variable) exerted by the water column by fixed in-situ pressure sensor and corrected to read zero at sea level

6.2.1 Web Services

BODC's parameter vocabulary can be accessed using web services.

A web service is a collection of protocols and standards used for exchanging data between applications or systems. Software applications written in various programming languages and running on various platforms can use Web services to exchange data over the Internet in a manner similar to inter-process communication on a single computer. This interoperability (e.g. between Java and Python, or Windows and Linux applications) is due to the use of open standards.

Further information is available from the BODC Web Services home page: http://www.bodc.ac.uk/products/web_services/

6.3 Glossary

- **Constituents** – The tide can be represented by the sum of a series of sine waves of determined frequency “harmonic constituents”. The parameters of each sine wave are called “harmonic constants”, and are the amplitude (half the height) of the wave and phase, or time of occurrence, of the maximum
- **Datum shift** – Improper maintenance operation, an accident, or even a natural phenomenon such as an earthquake may produce a sudden jump in the reference level
- **HAT (Highest Astronomical Tide)** – The height of the water at the highest possible theoretical tide
- **LAT (Lowest Astronomical Tide)** – The height of the water at the lowest possible theoretical tide
- **Mean sea level** – The mean value of sea level extracted from a suitably long series of data
- **Residual** – The difference between the tidal predictions and observations.
- **Spike** – A point in the data series which has an anomalous value outside of the surrounding range
- **Tidal analysis** – The analysis of an observed sea level record into a set of harmonic constants which can be used for tidal predictions
- **Timing error** – Disruption to the time channel caused by mechanical clock drift or a mistake in the digitising/processing of data

6.4 Screening procedure

The data are screened using BODC’s in-house software, Edserplo. The standard procedure at BODC for the quality control of SIMORC water level data includes, where possible:-

- Producing a tidal analysis and comparing M2, S2, N2, K1, O1 constituents with previous data series and adjacent sites
- Screening the series, looking for spikes, gaps, timing errors and datum shifts
- Screening the series with previous series from the same site
- Screening the series with neighbouring stations covering the same period
- Other parameters, such as sea temperature and atmospheric pressure, can be displayed at the same time to aid in quality control
- Checking the statistics produced, i.e. mean sea level, with those produced in previous years

Figure 11 – Example of datum shift caused by instrument malfunction

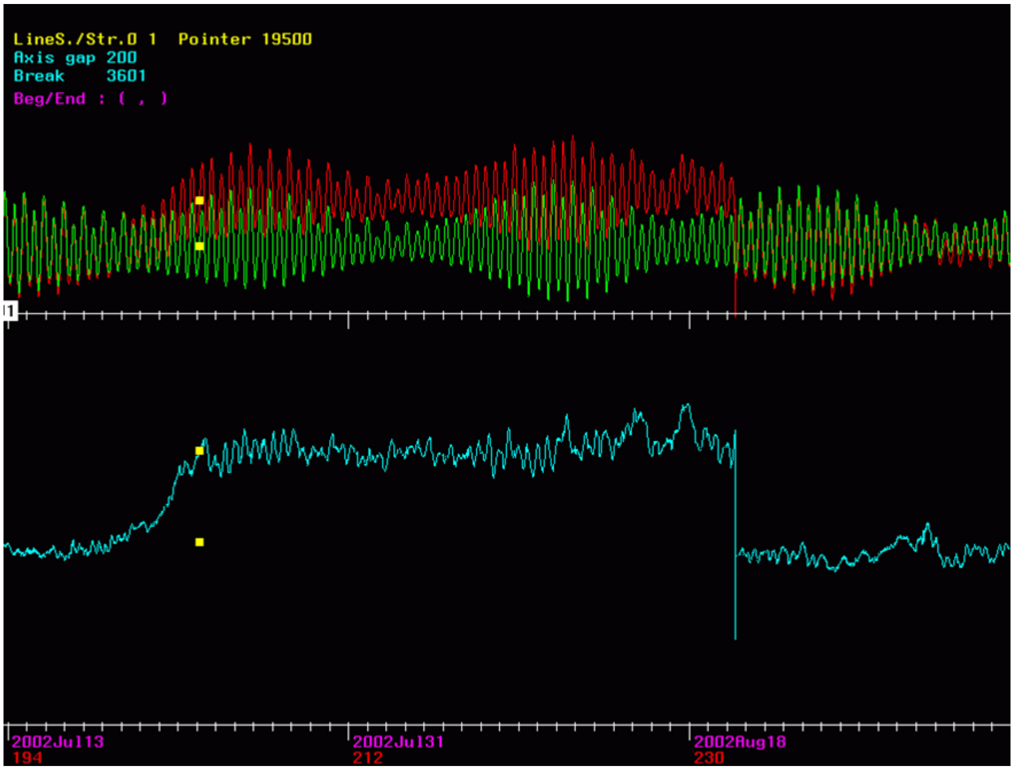


Figure 12 – Example of spikes

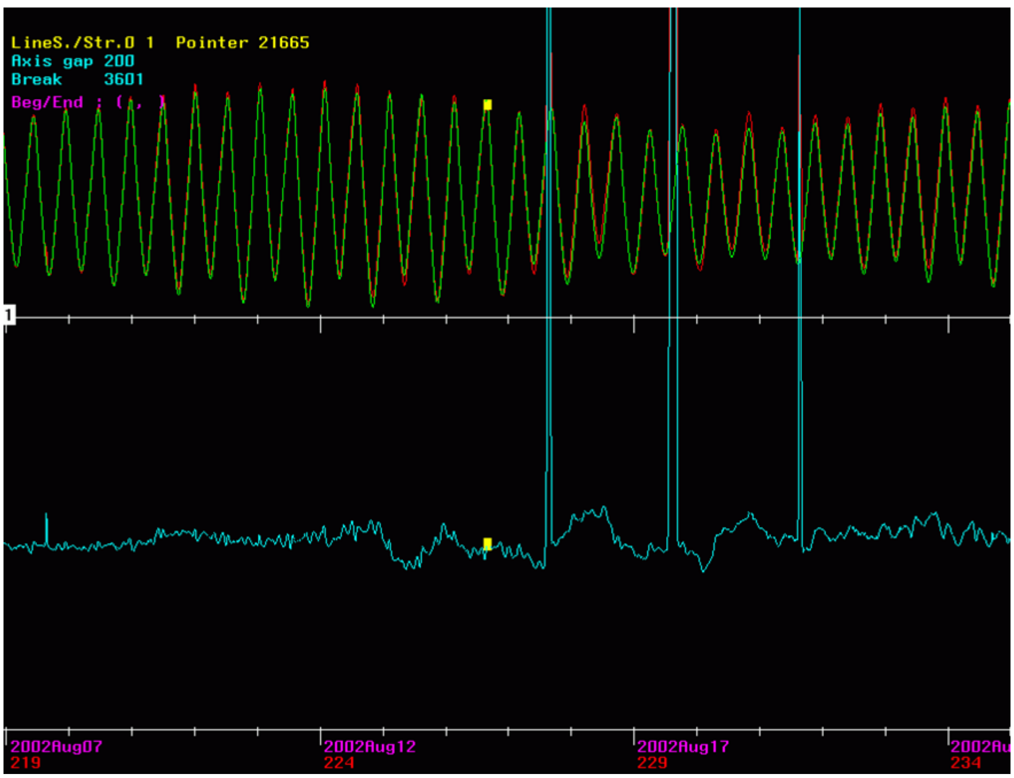
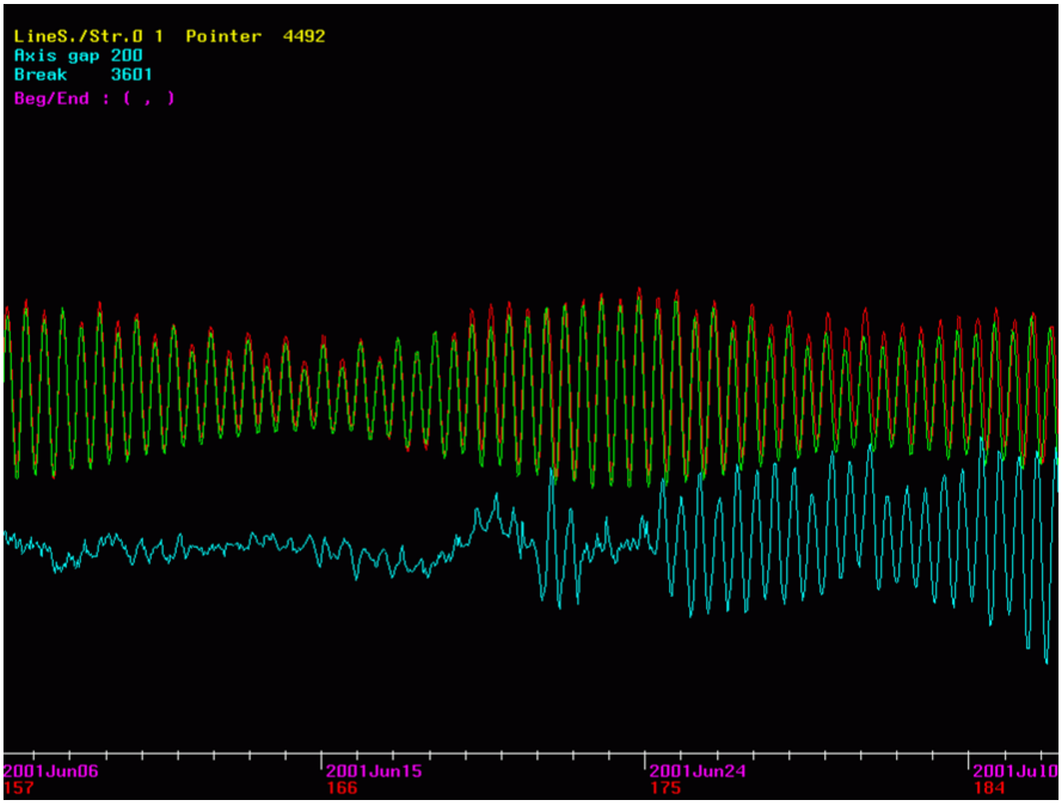


Figure 13 – Example of timing error caused by clock malfunction



(Legend: Tidal observations (m), Tidal predictions (m) and Residual (m))

Suspicious data points are flagged M and any timing errors or datum shifts are noted. No data values are changed. An N flag is assigned to those values that are null. The data quality is noted in accompanying documentation.

6.4.1 Tests on processed data

Listed below are a number of checks taken from the IOC/CEC's 'Manual of Quality Control Procedures for Validation of Oceanographic Data', 1993, which can be used to gauge whether data are realistic or not:

a) Mean level test

Where the water level has been recorded by a pressure recorder at depth, then the mean water level above the meter should correspond closely to the known depth of the meter below the surface. This test is carried out manually, and is used as an indicator that all is well.

b) Gross error limits

Maximum limits are (HAT-LAT range) + (1.2 x 100yr storm surge range). For UK waters, maximum storm surge range is 5m. Thus limits are defined by:
(LAT below ML - 3.0m) < WL < (HAT above ML + 3.0m)
where WL is the water level

c) Check limits

Check limits are defined by:
(LAT below NIL) < WL < (HAT above ML)

d) Rate of change check

The theoretical differences between consecutive samples h_1 and h_2 for various sampling rates Δt, assuming a semi-diurnal tidal period of 12.42, hours are given below:

Δt (min)	theoretical $\lvert h_1 - h_2 \rvert$	allowable $\lvert h_1 - h_2 \rvert$
0	0.0843 A	0.05 (HAT-LAT)
15	0.1264 A	0.08 (HAT-LAT)
20	0.1685 A	0.10 (HAT-LAT)
30	0.2523 A	0.15 (HAT-LAT)
60	0.5001 A	0.30 (HAT-LAT)

where A is tidal amplitude. The allowable difference given above, has been based on an amplitude of 0.5(HAT-LAT), with a 20% increase to allow for asymmetry in the tidal curve.
The flag is set against the second sample, h_2.

c) Stationarity check

Theoretically, for a sine or cosine curve a maximum number of two consecutive samples may have the same value (assuming no aliasing). However, in practice, the number of consecutive equal values depends on the tidal range and nature of the tidal curve at a site, the resolution of the tide gauge, and the sampling interval.
Suggested numbers of consecutive equal values allowed depending on the sampling interval are:

Δt (min)	Number of consecutive equal values allowed
10	12
15	8
20	6
30	4
60	2

This implies that stationarity of up to 2 hours is allowed, but any periods exceeding this are flagged.

f) Tidal range check

This is to assist in ensuring that no scale changes have occurred or that two data series have not been mismatched.
The tidal range from successive maxima (high waters) and minima (low waters) should lie between the minimum neap and the maximum spring (i.e. HAT-LAT) range.
Thus:

$\lvert h_{max} - h_{min} \rvert$
minimum neap range < or < HAT - LAT) range
$\lvert h_{min} - h_{max} \rvert$

where h_{max} and h_{min} are successive water level maxima and minima. Failure of this check causes a flag to be set against the second value.

h) Time of maxima and minima

For most cases, the time difference between successive h_{max} (high water) and h_{min} (low water) and between successive h_{min} and h_{max} should be between 4 ¼ and 8 ¼ hours.

$$\text{Thus } 4\tfrac{1}{4} \text{ hrs} < \begin{array}{c} |Th_{max} - Th_{min}| \\ \text{or} \\ |Th_{min} - Th_{max}| \end{array} < 8\tfrac{1}{4} \text{ hours}$$

where $Th_{ma}x$ and Th_{min} are the times of successive water level maxima and minima. Failure of this check causes a flag to be set against the second value.

6.5 Accompanying Documentation

At BODC we include a set of standard documentation with every data series. Using in-house 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 are as follows:

- 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.

- 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 that is used consistently for measurements over time.

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

- BODC Screening Document – This is a generic document linked to all datasets giving a brief summary of the screening procedure BODC undertakes 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.

7.0 Meteorological Data – Quality Control

7.1 Checklist of Metadata Required for Processing/QC/Documentation

The checklist and example information below shows the information used by BODC to ensure that the data are adequately described.

Owner Details	
Name of country responsible for data	e.g. UK
Name of organisation responsible for data	e.g. POL
Project Name (if applicable)	e.g. Coastal Observatory
Data Type (e.g. current, wave, sea level, met)	e.g. met observations
Mooring/Instrument Details	
Instrument category (e.g. current meter, wave recorder)	e.g. MAWS
Mooring/Rig Number	e.g. 1234
Instrument model and manufacturer	e.g. Young anemometer
Principle of measurement?	e.g.
Any instrument modifications?	e.g. height of the anemometer above sea level
Additional notes deployment	e.g. Any known data gaps / dropout problems
Additional notes on performance of mooring	e.g. high-frequency cut-off
Latitude of mooring (degrees)	e.g. 53.85°
Longitude of mooring (degrees)	e.g. -3.26°
Time zone	e.g. GMT/UTC
Site Area and Name of Site	e.g. Irish Sea, ABC site
Method of position fix	e.g. GPS
Water column Depth (m)	e.g. 352m
Sea Floor Depth Qualifier (e.g. echo sounder, GEBCO)	e.g. GEBCO
Timing Details	
Date and Time of Deployment (UTC)	e.g. 25/08/04, 10:00
Date and Time of start of usable data (UTC)	e.g. 25/08/04, 10:20
Date and Time of Recovery (UTC)	e.g. 14/04/05, 16:30
Date and Time of end of usable data (UTC)	e.g. 14/04/05, 16:10
Nominal time interval between successive data cycles in series (seconds)	e.g. 600s
Type of sampling (e.g. instantaneous, averaged)	Wind direction – vector averaged Air Temperature - instantaneous
Parameter Details	
Parameters measured and any definitions where the parameter is not obvious (See 3.3. BODC Parameter Dictionary below)	e.g. atmospheric pressure (CAPH%%) air temperature (CDTASS%%))
Data Processing Details	
Originator's Data Format	e.g. ASCII, .mat

Description of calibrations Description of any data processing that has occurred (manufacturers and in-house)	

7.2 BODC Parameter Dictionary codes

To get a comprehensive list of our parameter codes and their definitions, you can go to our online parameter dictionary at:
http://www.bodc.ac.uk/data/codes_and_formats/parameter_codes/bodc_para_dict.html

From here you can download the Parameter code list, the Parameter Group code list and the Units of Measurement code list, among others.

To aid your search we have included the most frequently used Parameter Group codes and Parameter codes for current meter data below:

Frequently Used Parameter Group Codes for Meteorological Data:

Parameter Group Code	Description
EWDA	Parameters expressing wind direction
EWSB	Wind speed
CDEW	Dewpoint of the atmosphere
CPRR	Precipitation rate
CDTA	Air temperature
CAPA	Barometric pressure

Frequently Used Parameter Codes for Meteorological Data:

Code	Description
CBTASS01	Temperature of the atmosphere by Acclaim barometer temperature sensor
CDTASS01	Temperature of the atmosphere by dry bulb thermometer
CDEWCV01	Dew point temperature of the atmosphere by computation from air temperature and relative humidity data
CDEWCVWD	Dew point temperature of the atmosphere by computation from wet and dry bulb thermometer data
CDEWZZ01	Dew point temperature of the atmosphere
CRELCV01	Relative humidity of the atmosphere by computation from wet and dry bulb thermometer data
CRELSS01	Relative humidity of the atmosphere by humidity sensor
CWETSS01	Wet bulb temperature of the atmosphere by psychrometer
CWETZZ01	Wet bulb temperature of the atmosphere
CPRRRG01	Precipitation rate (liquid water equivalent) in the atmosphere by in-situ rain gauge
CSLRR101	Downwelling vector irradiance as energy (solar (300-3000nm) wavelengths) in the atmosphere by pyranometer

EGSCSS01	Wind speed standard deviation (gust) in the atmosphere by in-situ anemometer
EGSCZZ01	Wind speed standard deviation (gust) in the atmosphere
EGTDSS01	Wind from direction (gust) in the atmosphere by in-situ anemometer
EGTDZZ01	Wind from direction (gust) in the atmosphere
EGTSSS01	Wind speed (gust) in the atmosphere by in-situ anemometer
EGTSZZ01	Wind speed (gust) in the atmosphere
ERWDSS01	Wind from direction (relative to moving platform) in the atmosphere by in-situ anemometer
RWDZZ01	Wind from direction (relative to moving platform) in the atmosphere
ERWSSS01	Wind speed (relative to moving platform) in the atmosphere by in-situ anemometer
EWDASS01	Wind from direction in the atmosphere by in-situ anemometer
EWSBSC01	Wind speed in the atmosphere by satellite scatterometer
CAPASS01	Pressure exerted by the atmosphere, measured by barometer

7.3 Glossary

- **Absolute wind speed –** average wind speed measured from a static base or corrected for platform velocity and heading
- **Absolute wind direction** – average wind direction measured from a static base or corrected for platform velocity and heading
- **Gust wind speed (1 minute) –** fastest instantaneous wind speed recorded over a 1 minute sampling interval
- **Gust wind speed (3 seconds)** - fastest instantaneous wind speed recorded over a 3 second sampling interval
- **Barometric pressure** – also known as atmospheric pressure. The weight of overlying air on the surface, at sea level, expressed in mbars
- **Air temperature** here this is taken to mean the temperature recorded on site, without any corrections for the altitude of measurement
- **Sea surface temperature** – taken to mean the temperature of the top 1 metre or surface mixed layer; this is not the same as the 'skin temperature' which is just the top few centimetres
- **Relative humidity –** ratio of the amount of water vapour in air to the maximum amount that could be held at that temperature, expressed as a percentage
- **PAR –** photosynthetically available radiation; the proportion of solar radiation which has a wavelength in the range 400 to 700nm
- **Solar irradiation/TIR** – electromagnetic radiation between wavelengths 300-3000 nm

7.4 Screening and QC Procedures – All Met Data

After reformatting using BODC's in house Transfer system, data will be screened on a graphics workstation. BODC's in-house software for quality controlling meteorological data comprises a visualisation tool called SERPLO, developed in response to the needs of BODC whose requirement involved the rapid inspection and non-destructive editing of large volumes of data. SERPLO allows the user to select specific data sets and to view them in various forms, to visually asses their quality. 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 might not otherwise be picked up if just viewing large columns of figures. 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. These flags do not change the data; they purely highlight potential problems with the data, allowing the end user to make the decision as to whether to use the data or not.

7.4.1 Time Series Plots

Using the time series plot all parameters can be visually 'screened' with the aim of looking for anomalous values. Often, related parameters, such as average wind speed and gust wind speed, are screened together to identify spurious values. This is because a sudden change in one of the parameters would lead you to expect a change in the other in agreement that it is a genuine event. The time series plots will be used to identify:

- Instrument failure (10 or more consecutive points of identical value)
- Definition of calm and appropriate flagging for wind speed and direction
- Wind direction either veering or backing during passage of a depression

7.4.2 Scatter Plots

Another useful tool provided by SERPLO is the ability to produce scatter plots of the wind vectors. These can show irregularities in the data, mainly as a result of mechanical malfunction. These plots can show larger than anticipated holes or 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.

Examples of scatter plots are shown below in figures 14 and 15:

Figure 14 – Example of satisfactory directional data from a data series

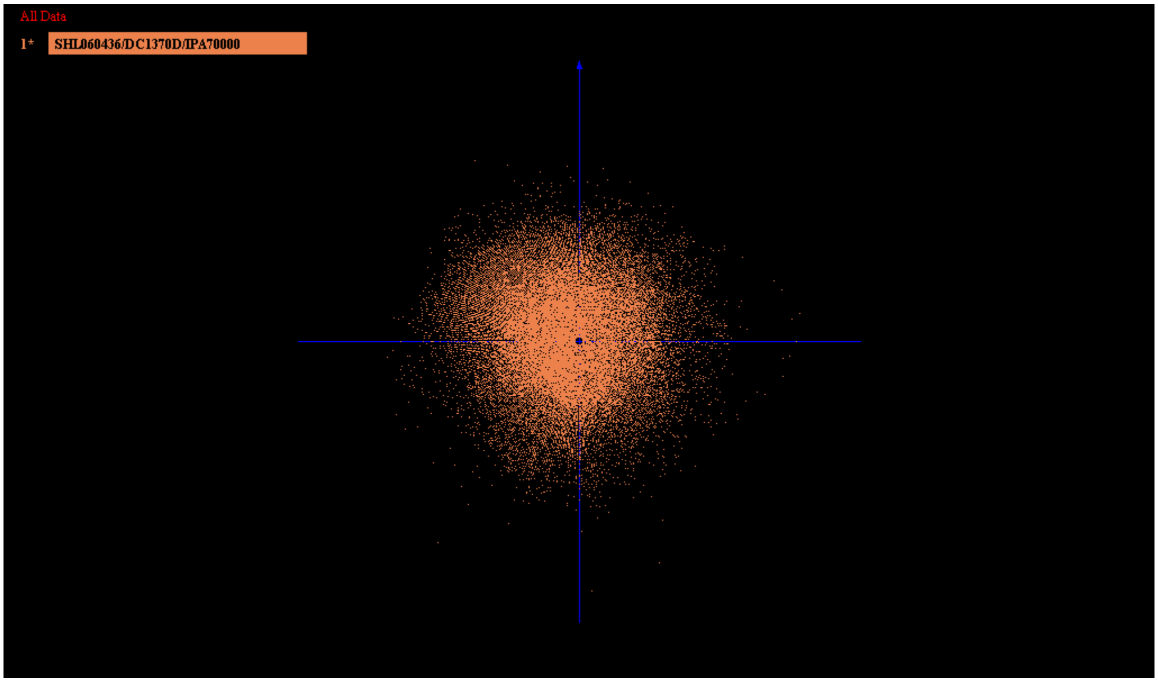


Figure 14 shows an example of a 'good' scatter plot. There seems to be an even array of wind directions indicating that the compass was not being hindered. However, Figure 15 below is an example of a record with suspect directions, as there are very few measurements between 80 and 100 degrees, or between 260 and 280 degrees.

Figure 15 - Example of suspicious directional data

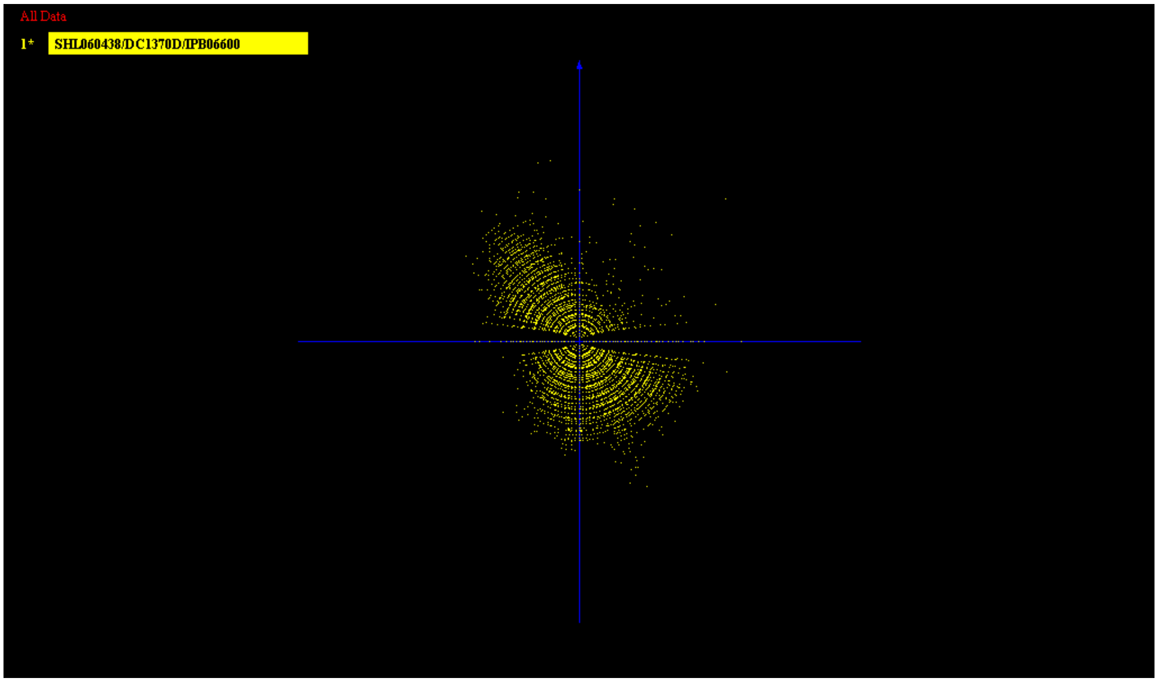


Figure 15 shows an example of a scatter plot that invites further investigation. In this instance, as a number of directions of 90 or 270 degrees were recorded, it was decided that there was no compass error, although possibly a shadowing effect of the placement of the anemometer.

Checks will be made for:

- gaps or preferential directions for wind speed and direction
- sudden changes in ratio of PAR/TIR e.g. for cloud cover

7.4.3 Problems Associated with Met Data

The main problems are expected to be constant values and possibly wandering means, which will be identified as above. Additionally, anemometer problems caused by wind speeds which are so low that the wind direction cannot be recorded will be identified.

7.5 Accompanying Documentation

At BODC we include a set of standard documentation with every data series. Using in-house 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 are as follows:

- Data Activity – this document describes the 'event'. In the case of met buoys or similar it generally gives the dates of deployment and 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 that 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.

- 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.

- BODC Screening Document – This is a generic document linked to all datasets giving a brief summary of the screening procedure BODC undertakes 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. For instruments such as anemometers, details of location, such as height above sea level should also be included.

8.0 Audit Procedures

To carry out an audit on a dataset that has been completely processed, a data scientist other than the processor is required to complete a series of final checks. This is done to give further confidence in the quality of the data. Any differences of opinion can be highlighted and looked at again, hopefully with a common solution being agreed upon.

8.1 Screening

This involves having a final look at the QXF files using our in-house visual screening software, Edserplo, to find any inconsistencies. The normal procedure is to do this channel by channel for all the data series from the same mooring using a common scale. Flags can be added or deleted where necessary in Edserplo before the files are banked and any changes noted.

8.2 Parameter Checks

The parameter codes assigned by BODC to data channels are then checked. The definition of each code is checked against the data in the dataset to ensure that the correct parameter has been assigned. A parameter set number is also assigned to individual data series, which defines which parameters are present for each series.

8.3 Documentation Checks

The documents written to accompany the data series are checked for errors. Common errors can include misquoted calibration details, mooring names and geographical positions for example.

8.4 Oracle Table Checks

The Oracle tables are checked by an in-house Matlab script designed to cross-reference the data header files against the metadata stored in Oracle and in documentation. This metadata includes information such as the start and end dates of data collection, latitudes and longitudes of moorings and depths of moorings to ensure no data-entry errors have been made. Another Matlab script checks the data files themselves to ensure anomalies such as timing errors and out-of-range and null values have been dealt with.

8.5 Checklists

A checklist is used to track and document the audit procedure plus a narrative that describes any problems encountered and how they were rectified.

9.0 Data Distribution and Delivery

Promotion and dissemination will be a continuous activity throughout the full project and it will be very much down to the partners to be pro-active in creating awareness among potential users of the new facility. The objective is to promote and disseminate the SIMORC facility in both the oil and gas industry and scientific communities, to achieve an increasing number of users and contributing parties.

IOC-IODE (The Intergovernmental Oceanographic Commission-International Oceanographic Data and information Exchange) has the primary role of disseminating information on the SIMORC facility to the scientific community worldwide, and the promotion of the SIMORC database and its possibilities to all potential users worldwide. The promotion and dissemination will be organised through a variety of methods including:

- SIMORC website which will be constantly updated – http://www.simorc.org
- Adding links to SIMORC on other websites including those of the Partners, the Advisory Board members and other Oceanographic Institutions e.g. IODE OceanPortal, Sea-Search. By increasing the number of websites that have the SIMORC link, the higher the search priority will be in search engines such as Google
- The production and distribution of posters and leaflets
- Presentations of the project in scientific meetings
- Presentations in other new project launches which are relevant e.g. SeaDataNet
- Advisory Board meetings and feedback to their user communities
- Ocean Teacher Training Tool, which is an IOC training tool. The SIMORC protocol will be included in the training curriculum once the SIMORC standards are available – http://oceanteacher.org
- Articles in scientific magazines

The data will have entries in the BODC International Current Meter Inventory. While the holdings will be present in the inventory (i.e. for search purposes only) access will not be gained to them through the BODC website. There will be instructions directing users to the SIMORC website, from where they will be able to download the data. Data held at BODC will therefore be restricted and will not be downloadable from the BODC website.

9.1 Data Formats

Data will be available from the SIMORC Data Delivery system in three formats initially. These are the BODC Request format (ASCII) and QXF(NetCDF). These are described below. Other formats may be added later to fulfill the requirements of SeaDataNet. ODV format (ASCII) and a specific SeaDataNet NetCDF format are likely to be included.

9.1.1 BODC ASCII Format

This is a generalised output format to handle most types of data held in the BODC databases. The following is an example of a file listed in this format together with a line by line explanation.

```
BODC Request Format Std. V1.0           Headers=  17 Data Cycles=    444           (a)
Series=      690343                     Produced: 03-Oct-2007                      (b)
Id                     6051_CURR Neth> Shell International Exploration and Pro>   (c)
053d5.1mN002d9.1mE                      start:19920723095300 end:19920728003800    (d)
Dep: floor   35.0 sensor    7.0     7.0 Nom. sample int.:  900.0 SEC               (e)
    3 Parameters included:                                                         (f)
Parameter f    P    Q Absent Data Value Minimum Value  Maximum Value       Units
ACYCAA01  Y   32   42             -1.00          1.00         444.00     Dmnless   (g)
Sequence number
LCDAEL01  Y   44   54             -1.00          1.00         358.00       deg T
Current direction (Eulerian) in the water column by in-situ current meter and c>
LCSAEL01  Y   56   66             -1.00          3.00          73.00        cm/s
Current speed (Eulerian) in the water column by in-situ current meter

Format Record                                                                      (h)
(10,19,1,1,%9d,1,1,%6.2f,1,1,%6.2f,1,1)



  Cycle      Date      Time    ACYCAA01:1  LCDAEL01:1  LCSAEL01:1                  (i)
Number    yyyy/mm/dd hh:mi:ssf          f           f           f
       1) 1992/07/23 09:53:00           1      339.00       12.00                  (j)
       2) 1992/07/23 10:08:00           2       20.00       17.00
       3) 1992/07/23 10:23:00           3       37.00       12.00
       4) 1992/07/23 10:38:00           4       60.00       12.00
       5) 1992/07/23 10:53:00           5       86.00       13.00
```

Notes

(a) The first record contains general information regarding the file. 'Std.' indicates Standard format and V1.0 indicates version 1.0 of the format. 'lrec' is the length (in bytes) to which the records are blank padded. Files may be produced without blank padding in which case no lrec attribute is included in the record. 'Headers' and 'Data Cycles' are counts of the number of header records and data cycles in the file. 'BODC QC' indicates that the data has been through BODC quality control procedures — this field is blank if this is not the case.

(b) Record two presents the BODC series reference number. If a series has not yet been allocated a BODC reference number this record will start with 'File:' followed by the full BODC file name. This record also indicates the date on which the output was produced (yyyy/mm/dd).

(c) Record three gives the data originator's identifier for the series, the source country and the source laboratory. If this information is not available the record will state 'Series header information not available' and the next two records will be blank.

(d) This record specifies one or two geographic positions. If a second position is given its purpose will be described in the accompanying documentation. Start date and end date (if available) are given in the format yyyymmddhhmiss (24 hour clock and UTC). If time is unavailable, hhmiss will be blank.

(e) This record gives the sea floor depth and the sensor depth. If a second (greater) sensor depth is given the two sensor depths specify the range of depths over which measurements were made. The second half of this record gives the nominal sampling interval and its units.

(f) This record and the following title record start the parameter section. There are two records per parameter present.

(g) The parameter information record gives

- the parameter code as defined in the BODC Parameter Dictionary
- whether the channel has been flagged with quality control indicators (Y/N)
- byte pointers (P and Q) to the start and end of the parameter within each data cycle record
- the absent data value
- minimum and maximum values of the parameter within the series
- parameter storage units

The next record gives the full description of the parameter. Units, parameter name and sampling method may be blank if series header information was not available.

(h) This line indicates the number of following records which together describe the Fortran format used to write each data cycle record.

(i) This and the next record are the data cycle title lines. 'f' indicates a flag channel.

(j) Data cycles are listed one per line. The first seven characters are always a data cycle count. One of the following quality control flags may appear against an individual data value (if the remark 'BODC QC' is present in record 1, then a blank flag indicates that the value is good).

<u>Multi Dimensional data</u>

Multi dimensional data such as ADCP data includes an extra section between the format record and data cycles records specifying values in the secondary dimension, e.g.

```
          DBINAA01
       1)       3.0
       2)       7.0
       3)      11.0

  Cycle       Date      Time    ACYCAA01:1  LCDAAP01:1  LCDAAP01:2  LCDAAP01:3
Number    yyyy/mm/dd hh:mi:ssf           f           f           f           f
       1) 1993/05/16 10:09:59           1      105.00      113.00      127.00
       2) 1993/05/16 10:20:00           2      103.00      114.00      125.00
       3) 1993/05/16 10:30:00           3      111.00      117.00      125.00
```

Each parameter will have a column for each value in the secondary dimension. This example has current direction for each of the three bin depths.

<u>Flag description</u>

	Good data
<	Below detection limit
>	In excess of quoted value
A	Taxonomic flag for affinis (aff.)
B	Beginning of CTD Down/Up Cast
C	Taxonomic flag for confer (cf.)
D	Thermometric depth

E	End of CTD Down/Up Cast
H	Extrapolated value
I	Taxonomic flag for single species (sp.)
K	Improbable value — unknown quality control source
L	Improbable value — originators quality control
M	Improbable value — BODC quality control
N	Null value
O	Improbable value — user quality control
P	Trace/calm
Q	Indeterminate
R	Replacement value
S	Estimated value
T	Interpolated value
U	Uncalibrated
W	Control value
X	Excessive difference

The following annotated listing of a header section will enable positions of any required fields to be calculated (dummy values have been entered for all possible fields).

9.1.2 QXF (aNetCDF) format

In addition to the ASCII format, data files are also supplied in BODC QXF format. QXF is BODC's primary datacycle storage format and is based on an internationally recognised platform-independent exchange format (NetCDF). Consequently, it may be considered as a viable data delivery format for some of our customers. QXF has the ability to handle multi-dimensional data such as ADCP data and spectral wave data.

NetCDF is a format used by and created for, the atmospheric research community in the USA. It is described in more detail in the section below. NetCDF interface software is freely available from and supported by the Unidata Program Center. This is managed by the University Corporation of Atmospheric Research (UCAR) on behalf of the US National Science Foundation. A toolbox allowing NetCDF files to be read into MATLAB may be obtained from the United States Geological Survey.

The information provided below assumes you have some prior knowledge of software programming.

Basic concepts

QXF is 'a format for datacycles'. Consequently, it does not include the metadata, even the fields that are essential for the effective use of the data. If data are delivered in QXF then BODC practice is to deliver the metadata separately either as a simple ASCII text file or an Extensible Markup Language (XML) file.

In NetCDF there are variable attributes and global attributes

- Variable attributes are used for storing such items as absent data values, formats and limits
- File or global attributes can be used to store metadata relating to the file as a whole

NetCDF operates with the notion of dimension. The primary dimension is the NetCDF's record dimension. This is normally time and is always present. The secondary dimension (which can be absent) contains information such as observing depths.

Each QXF file contains a header (a series of global attributes) and a series of QXF channels. These channels generally consist of a data variable (integer or single-precision floating point) plus a 1 byte quality control flag channel. The flag channel will not be encountered in the following circumstances

- Channels that are attached to the 'secondary' dimension alone (zero rank — see below)
- BODC standard date channel (the flag is carried by the separate time channel)

The data variables are named using the 8 byte alphanumeric codes defined in the BODC Parameter Dictionary. The associate flag variables are named using the data variable names prefixed by an 'F'.

<u>QXF header</u>

The QXF header consists of NetCDF global attributes. The complete list, including data types, for QXF 1.0 is

QXFVER	Long	QXF version. Two element array set at (1,0)
SERIDN	Char*40	Internal file identifier. Series name
GOODFL	Char*1	Value of good flag (always set blank)
NULLFL	Char*1	Value of null flag (always set to 'N')
STOROP	Long	Storage option flag (always set zero)
NSCHAN	Long	Number of QXF channels
NSOFDA	Long	Number of storage locations used for the file (not bytes)
HEADSZ	Long	Header size expressed in locations
CYCLSZ	Long	Cycle size expressed in locations

Most of these fields are either designed for use by BODC systems or are to allow non-standard usage of the format within the data centre. Consequently, they are of little interest to external users. However, it is worth checking that the format version hasn't changed and that the Internal File Identifier is as expected. Files that have been completely processed by the BODC system will have this set to BSRnnnnnnnn, where 'nnnnnnnn' is the BODC Series Reference for the data.

QXF rank

Rank is a mathematical term used to describe the number of dimensions possessed by an array. The term is used here in a comparable sense for QXF channels.

- Zero rank — Channels attached to ‘secondary’ dimension alone
- Rank 1 — Channels attached to ‘primary’ dimension alone
- Rank 2 — Channels attached to both dimensions

QXF channel types

Three types of QXF channel need to be known about by the user:

1. **'Rank Zero' channels** — These are only found in data that include 2-dimensional netCDF data variables. The rank zero channels are associated with the NetCDF secondary dimension. Often there will only be one rank zero channel (e.g. bin depth for a moored ADCP), but more than one is a possibility (e.g. cell latitude and cell longitude for OSCR data). Rank zero channels do not carry a flag variable.

2. **Date channel** — These consist of a 1-dimensional long integer data variable named 'AADYAA01', a 1-dimensional single precision floating point data variable named 'AAFDZZ01' and a 1 byte character flag channel named 'FAAFDZZ01'. Date channels may not always be included as some data types (e.g. CTD profiles) may have other parameters such as depth as the primary independent variable.

   Variable 'AADYAA01' contains the date expressed as a 'Loch Daynumber'. This is the number of whole days elapsed since 00:00 on 01/01/1760. Open source conversion functions to change Loch Daynumbers into 'year, month and day' are available from BODC in Fortran and Pascal or inbuilt Matlab functions may be used. It may help users to know that the Loch Daynumber for 'C-Zero' (01/01/1970) is 76701 and the Loch Daynumber for 01/01/2000 is 87658.

   Variable 'AAFDZZ01' contains the time expressed as a day fraction (e.g. 06:00 corresponds to 0.25, 18:00 corresponds to 0.75, etc.).

   Variable 'FAAFDZZ01' is used to signify quality aspects of the whole datacycle. If it has been set non-blank, it does not mean that there is a problem with the time. Rather that the quality information applies to all parameters in the datacycle. Users should not encounter non-blank datacycle flags in fully processed data as they are only used to indicate datacycles that are to be deleted during BODC processing.

3. **Data channel** — These contain a long-integer or single precision floating point data variable that may be either one or two-dimensional. The variable is named using the appropriate BODC parameter dictionary 8-byte code. Each data variable is accompanied by a single-byte flag variable of equal dimension. The full list of flag definitions is given below. In practice, many of these flags have only been used on

rare occasions and are unlikely to be encountered. The flags L and M apply to data regarded as suspect.

Variable attributes

Each netCDF variable in a QXF file carries a set of variable attributes (sometimes termed channel attributes). These are named using the variable name suffixed by 'MIN', 'MAX', 'LFM', and 'ABS' (e.g. TEMPPR01.MAX). Data variables carry all four variable attributes. Flag variables only carry the first two.

The attributes are defined as follows

MIN — Minimum value stored in the variable array. For data variables, it includes all values except for absent data values, whatever the status of the accompanying flag.

MAX — Maximum value stored in the variable array. For data variables, it includes all values except for absent data values, whatever the status of the accompanying flag.

LFM — Information on a suitable output format for the data held in the array. In the simple case the LFM (logical format element) specifies the number of places before (BEF) and after (AFT) the decimal point.

- LFM = 100*BEF + AFT

Thus 100 would equate to F2.0 in Fortran for a float or real variable and I1 for an integer. In C the equivalent would be %2f and %1d. Note that space for a leading sign must be met by BEF.

Scientific notation, that is the use of mantissa and exponent, is met by inverting LFM's sign and setting BEF to 1. The number of significant digits is then 1 + AFT and the form is

- sd.<m.. ..m>Esxx

where <m.. ..m> is the part of the mantissa following the decimal point, 's's are the signs, and 'xx' is the exponent. LFM =-105 is equivalent to %12.5e (C), 1pe12.5 (Fortran). -0.0825 in this format appears as

- -8.25000e-02

Note. The space for the mantissa's sign is automatically allocated and it is not possible to define the case of the letter 'e' introducing the exponent. The presentation is chosen to coincide with that used in C's standard library.

ABS — This is the absent data value used in the variable array. This value (and no other) will be accompanied by a flag variable element set to 'N'.

10.0 References

- BODC (Version 3.12, March 2006) NODB Tables and Fields: A Guide to the Tables and Fields of the BODC National Oceanographic Database. BODC/STND/12.

- IOC/CEC, 1993, 'Manual of Quality Control Procedures for Validation of Oceanographic Data'

- Rickards, L.J., 1989, 'The UK National Oceanographic Data Centre – Current Meter Data: Quality Control and Banking', International Council for the Exploration Council Cooperative Research Report: Current Meter Data Quality, p.37-43

- WMO Publication number 702, 1998, 2nd edition, 'Guide to Wave Analysis and Forecasting'

- WOCE Data Products Committee. 2002. WOCE Global Data, Version 3.0, WOCE International Project Office, WOCE Report No. 180/02, Southampton, UK.

Annex 1: Extract from NODB Tables and Fields: A Guide to the Tables and Fields of the BODC National Oceanographic Database - "Series Header Information" section

ZESH - SERIES HEADER

ISHREF 'Series Reference No.' *(I8) - Mandatory*

Unique number assigned by BODC to identify the data series; numbers are allocated in sequence according to the appropriate Modulus 11-Check Algorithm (see Appendix C).

CWARNS 'Warning Document Flag' *(C1) - Null value blank*

Set to 'C' if the series is linked to a Problem Report Document (ZOND.CNDCAT set to 'C'). In exceptional circumstances, CWARNS may be set to 'W' to indicate that BODC consider the series unfit for **ANY** purpose. This should be considered as a non-destructive alternative to deleting the series from the databank. Otherwise, CWARNS is left blank.

IRSREF 'Restriction Record Reference No.' - see ZORS *(I8) - Mandatory*

This field identifies the confidentiality reference. It is made mandatory because it also covers related topics such as whether the series originator is to be informed about the release of the series to third parties, etc. Series which are free of any confidentiality restrictions are linked to IRSREF number 1310. Numbers are allocated in sequence according to the appropriate Modulus 11-Check Algorithm (see Appendix C).

Organisation Originating the Data Series

This should reflect the organisation responsible for the original creation of the data series and is expressed in terms of both a country code and an organisation code as follows:

CSHCTY 'Code for Country of Series Originator' *(C2) - Mandatory*

Intergovernmental Oceanographic Commission (IOC) Country Code as defined in BODC Code Table C18.

CSHORG 'Code for Organisation of Series Originator' *(C3) - Mandatory*

BODC code, assigned within country, identifying the organisation as defined in BODC Code Table C75.

COPCAT 'Primary Oceanographic Data Category' *(C2) - Mandatory*

BODC code describing the primary category of oceanographic data contained within the data series. Coded as in BODC Code Table C09.

COSCAT 'Secondary Oceanographic Data Category' *(C2) - Mandatory*

Should the type of data contained in the data series be fully covered by the entry in COPCAT then this field will be set to 'ZZ'. Otherwise it will be set to the secondary category of oceanographic data contained within the data series coded as in BODC Code Table C09.

CMOUNT 'Instrument Mounting Category' *(C2) - Mandatory*

BODC code describing the type of platform on which the sensors providing data for the series were mounted. Coded as in BODC Code Table C10.

CINCAT 'Instrument Category' *(C2) - Mandatory*

BODC code describing the type of instrumentation or methodology employed for the collection of the data contained in the data series. Coded as in BODC Code Table C11.

NOTE: It is inherent in the coding of the above four fields that each data series stored in the database will generally relate to a specific instrument package recording a specific type of oceanographic data. However, it is intended that if 'hybrid' data series are specifically required intact (e.g. data buoys), the coding of the above four fields should not prevent this.

Geographic Position of the Data Series

The use of the following four fields for describing the geographic position of the series is defined in field CSPOSD. Optionally they may be used to describe an accepted single position for the series as a whole, or the positions at the start and end of the series, or a geographic box within which all data in the series was collected.

FSLATA 'Latitude/Start Lat./Southern Lat. of the Series in degrees' *(F9.5) - Mandatory*

Negative values are south of the Equator.

FSLONA 'Longitude/Start Long./Western Long. of the Series in degrees'(F10.5) - Mandatory

Negative values are west of Greenwich.

FSLATB 'End Latitude/Northern Lat. of the Series in degrees' *(F9.5) - Can be null*

Negative values are south of the Equator.

FSLONB 'End Longitude/Eastern Long. of the Series in degrees' *(F10.5) - Can be null*

Negative values are west of Greenwich.

CSPOSD 'Series Position Fields Definition' *(C1) - Mandatory*

Set to 'F' if FSLATA, FSLONA contain the accepted latitude, longitude of the series as a whole in which case FSLATB and FSLONB are both set to null. Set to 'S' if FSLATA, FSLONA contain the latitude, longitude of the position at the start of the series and FSLATB, FSLONB the latitude, longitude at the end of the series, i.e. positions of the first and last datacycles in the series. Set to 'R' if FSLATA, FSLATB contain the southern and western latitude limits within which data in the series was collected and FSLONA, FSLONB contain the northern and eastern longitude limits of the series. (Note: Further entries may be defined in future: check the Code Table). Coded as in BODC Code Table C14.

CSPOSU 'Series Positional Uncertainty' *(C1) - Mandatory*

This is a code that describes positional uncertainty, coded as in BODC Code Table C07. If CSPOSD = 'F' it describes the uncertainty in FSLATA, FSLONA in terms of positional drift, positional scatter of sensors or the estimated inaccuracy in the measurement of position whichever is the greater or, if appropriate, a suitable combination of all three. Otherwise it is only entered where position is included at the datacycle level in which case it refers to the average estimated inaccuracy (in absolute terms) of the individual datacycle positions. If set to 'G' then FSLATA is set to -90, FSLONA to -180, FSLATB to

90 and FSLONB to 180. (Note: Further entries may be defined in future: check the Code Table).

Depth of the Data Series

FFLOOR 'Sea Floor Depth' *(F9.2) - Can be null*

Sea floor depth below sea level at the position FSLATA, FSLONA, expressed in metres. Normally only entered for series with data collected at a fixed location.

CFLDPQ 'Sea Floor Depth Qualifier' *(C1) - Null if FFLOOR null*

BODC code defining the datum from which the depth FFLOOR is measured. Coded as in BODC Code Table C06.

Depth Horizons sampled in the data series:

The following four fields relate to the depth range covered by the sensors providing data for the series. The exact definition of the fields FDEPTH and FMXDEP depend on the sensor configuration with respect to depth and on whether the sensors remain at a fixed depth. Their definition is determined by the entry made in field CSDEPD. It should be noted that these fields are used mainly for retrieval selection and that more accurate depth information may be included in Narrative Documents pertaining to the series or indeed in the Datacycle Records. By convention depth is measured positively downwards from the sea surface, and sensor heights above sea surface are expressed as negative depths.

FDEPTH 'Depth/Minimum Depth for Series' *(F9.2) - Can be null*

Where all sensors are fixed at the same depth with minimal depth spread for the duration of the series this field contains the accepted depth of the sensors below sea level. However, where the sensors are distributed with respect to depth, the data are collected from a series of depth bins or where sensor depths vary during the series, this field contains the minimum depth horizon (or bin depth) below sea level sampled during the series. In both cases the depth is expressed in metres.

FMXDEP 'Maximum Depth for Series' *(F9.2) - Can be null*

Where all sensors are fixed at the same depth with minimal depth spread for the duration of the series this field is set to same value as FDEPTH. However, where the sensors are distributed with respect to depth, the data are collected from a series of depth bins or where sensor depths vary during the series, this field contains the maximum depth horizon (or bin depth) below sea level sampled during the series. In both cases the depth is expressed in metres.

CSDEPD 'Series Depth Field Definition' *(C1) - Null if FDEPTH and FMXDEP null*

BODC code set to describe the variation in depth of the sensors with respect to each other and whether they are fixed for the duration of the series. Coded as in BODC Code Table C24.

CSDEPQ 'Series Depth Field Qualifier' *(C1) - Null if FDEPTH and FMXDEP null*

BODC code defining the datum from which the depths FDEPTH and FMXDEP are measured. Coded as in BODC Code Table C06.

Date/Time at the Start/End of the Data Series

This should contain the date/time, expressed in GMT of the earliest/latest valid datacycle recorded for the series, that is stored on the database. The precision with which the time can be estimated will normally be to the nearest minute, but in some cases it will be no better than one hour or even a day.

TSHBGN 'Series Start' *(DT) - Mandatory*

TSHEND 'Series End' *(DT) - Can be null*

CTMPRC 'Precision of Time Specification' *(C1) - Mandatory*

This applies to both TSHBGN and TSHEND. Coded as in BODC Code Table C28.

Cycle Interval

This field is used in retrieval for giving an indication of the characteristic interval between datacycles, but does not necessarily imply that data is collected precisely at the specified interval or that complete data coverage is available at the interval. It is however only meaningful where a considerable degree of regularity is present within the series.

FINTRV 'Nominal Interval between Datacycles' *(F8.2) - Can be null*

Expressed in units as defined in CIUNIT. Where no degree of regularity is present it should be set null.

CIUNIT 'Units of Nominal Interval between Datacycles' *(C3) - Null if INTERV null*

BODC code for describing the units used for the FINTRV entry as defined in BODC Code Table C08.

IPSREF 'Parameter Set Reference No.' - see ZEPS *(I6) - Mandatory*

CHECKS 'Series Quality Class' *(C1) - Mandatory*

BODC Code to describe the Quality Class assigned to the series. 'A' is used to show that the series is banked. Otherwise, CHECKS it is set to 'X'. Any other value defined in the C21 code table should not be used. Note that only records with CHECKS="X" may be updated from the BODCCONV user id.

CILOAD 'Load No.' - see ZILO *(C6) Mandatory*

TSHMOD 'Modification Date/time' *(DT) - Initially null*

Date/time of last modification effected on the record.

Annex 2: Examples of accompanying documentation

Project Document Example:

North Sea Project

The North Sea Project (NSP) was the first Marine Sciences Community Research project of the Natural Environment Research Council (NERC). It evolved from a NERC review of shelf sea research, which identified the need for a concerted multidisciplinary study of circulation, transport and production.

The ultimate aim of the NERC North Sea Project was the development of a suite of prognostic water quality models to aid management of the North Sea. To progress towards water quality models, three intermediate objectives were pursued in parallel:

- Production of a 3-D transport model for any conservative passive constituent, incorporating improved representations of the necessary physics - hydrodynamics and dispersion;
- Identifying and quantifying non-conservative processes - sources and sinks determining the cycling and fate of individual constituents;
- Defining a complete seasonal cycle as a database for all the observational studies needed to formulate, drive and test models.

Proudman Oceanographic Laboratory hosted the project, which involved over 200 scientists and support staff from NERC and other Government funded laboratories, as well as seven universities and polytechnics.

The project ran from 1987 to 1992, with marine field data collection between April 1988 and October 1989. Fifteen survey cruises (Table 1), each lasting 12 days and following the same track, were repeated monthly. The track selected covered the summer-stratified waters of the north and the homogeneous waters in the Southern Bight in about equal lengths together with their separating frontal band from Flamborough head to Dogger Bank, the Friesian Islands and the German Bight. Mooring stations were maintained at six sites for the duration of the project.

The data collected during the observational phase of the North Sea Project comprised one of the most detailed sets of observations ever undertaken in any shallow shelf sea at that time.

Fixed Station Document Example:

North Sea Project Dover Strait Site A
Site A was one of three stations, located between Dungeness and Cap Griz-Nez, where moorings were deployed during the North Sea Project Dover Strait study. From May 1990 to July 1991 an experiment was conducted to determine the flux of contaminants passing from the English Channel into the North Sea.

The site was characterised by strong rectilinear currents, with water depth of approximately 30m, ensuring a well-mixed water column throughout the year. Moored instruments were deployed at monthly intervals for the full period at Site A.

The rigs deployed at all sites in the Dover Strait study lie within a box bounded by co-ordinates 50.778N 1.231E at the southwest corner and 50.954N 1.534E at the northeast corner. Magnetic variation at this site was 3.6° west.

The thirteen pressure gauge deployments at this site did not maintain a common datum.

Site A deployment history is summarised below:

Rig ID	Meter type	Meter height	Start date	Data return (days)	Comment
00424	ADCP WR	0.5m 0.5m	22/05/90 22/05/90	8.7 8.7	
00425	TR CM	5.5m 4.0m	22/05/90 22/05/90	0.0 7.6	No data N/A
00434	ADCP WR	0.5m 0.5m	15/06/90 15/06/90	14.2 14.0	
00436	ADCP WR	0.5m 0.5m	18/07/90 18/07/90	34.1 34.2	
00438	ADCP WR	0.5m 0.5m	04/09/90 04/09/90	19.5 19.5	
00440	ADCP WR	0.5m 0.5m	24/09/90 24/09/90	23.7 23.7	
00454	ADCP WR	0.5m 0.5m	18/10/90 18/10/90	33.1 33.1	
00456	ADCP WR	0.5m 0.5m	20/11/90 20/11/90	27.9 27.9	
00466	ADCP WR	0.5m 0.5m	18/12/90 18/12/90	0.0 35.9	No data N/A
00468	ADCP WR	0.5m 0.5m	23/01/91 23/01/91	0.0 33.0	No data N/A

00470	ADCP WR	0.5m 0.5m	24/02/91 24/02/91	14.1 44.0	
00472	ADCP WR	0.5m 0.5m	10/04/91 10/04/91	0.0 33.0	No data N/A
00474	ADCP WR	0.5m 0.5m	13/05/91 13/05/91	26.3 28.9	
00476	ADCP WR	0.5m 0.5m	11/06/91 11/06/91	26.9 26.9	

ADCP = Acoustic Doppler Current Profiler
WR = Water-level Recorder
TR = Transmissometer
CM = Current Meter (Aanderaa or S4)

Data Activity Document Example:

North Sea Project POLRIG#00456
This rig was deployed as part of the NERC North Sea Project Process Study in the Dover strait at site A.

The mooring comprised of two instruments, an Acoustic Doppler Current Profiler (ADCP) and an Aanderaa WLR. These were mounted on a low profile frame positioned on the seabed. The instruments were located 0.5m above the sea floor.

Rig position	50° 56.59' N, 001° 16.40' E
Water depth	30m

Instrument	**Parameters measured**
WR1038	Pressure and temperature
DP0010	Current speed and direction

Data Processing Document Example:

North Sea Project Dover Straits Moorings ADCP data processing

The 1MHz ADCP deployed during the North Sea Project Dover Straits Study at mooring sites A, B and C were configured as illustrated below. During the study, compasses consistently gave directions of flow different from those produced by Ocean Surface Current Radar (OSCR) and previous current meter records (Prandle et al 1993). A frame angle correction was subsequently applied to each data series (employing methods as described in the instrument description) and the data from the top (near surface) bins were excluded, as necessary, when they are not in the water.

ADCP set up details for Rigs 424, 428, 434, 436, 438, 440, 454 and 456

Sample period	10 minutes
No. of bins (cells)	24
No. of pings/ensemble	180
First bin height	3.9m
Bin separation	1.4m
Bin heights (1-24)	3.9m, 5.3m, 6.7m, 8.1m, 9.5m, 10.9m, 12.3m, 13.7m, 15.1m, 16.5m, 17.9m, 19.3m, 20.7m, 22.1m, 23.5m, 24.9m, 26.3m, 27.7m, 29.1m, 30.5m, 31.9m, 33.3m, 34.7m, 36.1m

ADCP set up details for Rigs 470, 474 and 476

Sample period	10 minutes
No. of bins (cells)	16
No. of pings/ensemble	275
First bin height	3.9m
Bin separation	1.4m
Bin heights (1-16)	3.9m, 5.3m, 6.7m, 8.1m, 9.5m, 10.9m, 12.3m, 13.7m, 15.1m, 16.5m, 17.9m, 19.3m, 20.7m, 22.1m, 23.5m, 24.9m,

Data processing at BODC

The data were converted from ASCII format into BODC QXF, a subset of NetCDF. The data where then visually inspected, for quality control purposes, using in house software.

References

Prandle, D., Loch, S.G. and Player, R. 1993. Tidal Flow through the Straits of Dover. Journal of Physical Oceanography, Vol. 23. No. 1, January 1993

Narrative Document Example:

General Data Screening carried out by BODC
BODC screen both the series header qualifying information and the parameter values in the data cycles themselves.

Header information is inspected for:

- Irregularities such as infeasible values
- Inconsistencies between related information, for example:
    - Times for instrument deployment and for start/end of data series
    - Length of record and the number of data cycles/cycle interval
    - Parameters expected and the parameters actually present in the data cycles
- Originator's comments on meter/mooring performance and data quality

Documents are written by BODC highlighting irregularities which cannot be resolved.

Data cycles are inspected using time or depth series plots of all parameters. Currents are additionally inspected using vector scatter plots and time series plots of North and East velocity components. These presentations undergo intrinsic and extrinsic screening to detect infeasible values within the data cycles themselves and inconsistencies as seen when comparing characteristics of adjacent data sets displaced with respect to depth, position or time. Values suspected of being of non-oceanographic origin may be tagged with the BODC flag denoting suspect value; the data values will not be altered.

The following types of irregularity, each relying on visual detection in the plot, are amongst those which may be flagged as suspect:

- Spurious data at the start or end of the record.
- Obvious spikes occurring in periods free from meteorological disturbance.
- A sequence of constant values in consecutive data cycles.

If a large percentage of the data is affected by irregularities then a Problem Report will be written rather than flagging the individual suspect values. Problem Reports are also used to highlight irregularities seen in the graphical data presentations.

Inconsistencies between the characteristics of the data set and those of its neighbours are sought and, where necessary, documented. This covers inconsistencies such as the following:

- Maximum and minimum values of parameters (spikes excluded).
- The occurrence of meteorological events.

This intrinsic and extrinsic screening of the parameter values seeks to confirm the qualifying information and the source laboratory's comments on the series. In screening and collating information, every care is taken to ensure that errors of BODC making are not introduced.

Instrument Document Example:

Aanderaa Recording Current Meter Model 7/8

Manufacturer's specifications: recording unit height 49.5cm (RCM8 52.0cm), diameter 12.8cm, vane size 48.5x50.0cm. Meter is designed for depths down to 2000m (RCM8 6000m). It incorporates a spindle which is shackled to the mooring line. The meter is attached to the spindle through a gimbal mounting which permits a maximum 27° deviation of the spindle from the vertical, the meter still remaining horizontal.

Meter comprises:

1. Paddle wheel rotor magnetically coupled to an electronic counter
2. Vane, which aligns instrument with current flow, has a balance weight ensuring static balance and tail fins to ensure dynamic balance in flows up to 250cm/s.
3. Magnetic compass (needle is clamped to potentiometer ring) - direction recorded with 0.35° resolution, 5° accuracy for speeds 5 to 100cm/s, 7.5° accuracy for remaining speeds within 2.5 to 200cm/s range.
4. Quartz clock, accuracy better than 2 sec/day within temperature range 0 to 20°C.
5. Thermistor (temperature sensor), standard range -2.46 to 21.48°C (max on high range 36.04°C), accuracy 0.05°C, resolution 0.1 per cent of range, 63 per cent response time 12sec.
6. Inductive cell conductivity sensor (optional), range 0 to 70mmho/cm standard resolution 0.1 per cent of range.
7. Silicon piezoresistive bridge, standard range 0 to 3000 psi (RCM8 to 9000 psi), resolution 0.1% of range.
8. Self balancing potentiometer which converts the output from each sensor into a 10 bit binary number for storage on magnetic tape.
9. Associated electronics.

A built-in clock triggers the instrument at preset intervals and up to six channels are sampled in sequence. Channel 1 is a fixed reference reading for control purposes and data identification. Channels 2, 3 and 4 represent measurement of temperature, conductivity and pressure. Channels 5 and 6 represent the VECTOR AVERAGED current speed and direction since the previous triggering of the instrument. The number of rotor revolutions and the direction is sampled every 12 seconds and broken into North and East components. Successive components are added and recorded as speed and direction. For recording intervals longer than 10 minutes, speed and direction are sampled 1/50th of recording interval.

It has become common practice in some laboratories to deploy these meters as temperature and conductivity loggers without current measuring capabilities.

Problem Report Example:

Problem Report

A visual examination of the data series indicates a serious problem with the instrument as it is showing bottom currents significantly stronger than surface currents.

The contributor has explained the instrument malfunction with the following statement:

A gradual reduction in current speed is apparent, with increasing distance from the transducer. This had been noticed on earlier experiments with the 250KHz ADCP, but is a greater problem with the 1 MHz ADCP measuring 24 bins. The ADCP setup was changed to decrease the problem with only limited success and continued loss of signal. A correction, using backscatter strength also recorded by the ADCP, was investigated to no avail.

Restriction Document Example:

Public Domain Data

This series has no specific confidentiality restrictions. However, the Intellectual Property Rights of the data originator must be respected, including acknowledgment in any publication or other output resulting from usage of the data.