SUGGESTED MODIFICATIONS TO THE ARGO BUFR FORMATTED MESSAGES

Background

Action 12 from the 14th Argo Steering Team meeting (March 2013) requested to 'develop BUFR templates for real time data distribution on the GTS'. Specifically this was aimed at modifying the existing BUFR template for Argo data to allow (i) un-pumped near surface temperature measurements, (ii) dissolved oxygen and (iii) bio-geochemical variables to be included; as the present template can only represent temperature and salinity profiles. As the Argo NetCDF format develops to allow for the inclusion of these additional measurements it is important that the Argo BUFR evolves to provide the same capability.

Outline of the BUFR format

BUFR (Binary Universal Form for the Representation of data) is the format that the WMO (World Meteorological Organization) has adopted for the exchange of time-critical data on the GTS (Global Telecommunications System). Note that from November 2014 the use of FM-64 TESAC to disseminate real-time Argo data should cease and profiles only disseminated on the GTS in BUFR. Documentation on the BUFR format is given at http://www.wmo.int/pages/prog/www/WMOCodes.html.

Appendix A outlines the format of the present Argo BUFR message, as distributed on the WMO GTS, and Appendix B gives the specific template used for standard temperature and salinity profiles from Argo floats.

The present Argo template is based on BUFR Master Table 0 which is used operationally by the international meteorological community. However, as it is fully self-describing, the BUFR approach is valid for any type of data and BUFR Master Table 10 has been developed for oceanographic data, where the governance of Master Table 10 comes under the WMO-IOC Joint Commission for Oceanography and Marine Meteorology (JCOMM). However, at the present time no BUFR templates have been developed using Master Table 10 and it has not been used. A key point to note is that while Master Tables 0 and 10 have some overlap in terms of the variables that can be represented, their descriptors are not inter-changeable. Hence an individual BUFR message must comply with either Master Table 0 or Master Table 10, but cannot comply with both.

At present Master Table 0 includes descriptors for depth (and/or pressure), temperature, salinity and dissolved oxygen, whereas Master Table 10 would allow for a wider range of oceanographic variables (chlorophyll-a, pH, nitrate content etc).

One of the requirements for changes to the Argo BUFR messages is to allow for the inclusion of additional bio-geochemical variables, as it is envisaged in the future bio-geochemical data will be used to initialise, constrain and validate operational real-time

ecosystem models. This implies a need to deliver these additional variables in real-time (or near real-time) on the GTS (as well as being available via the GDACs in NetCDF). For all non time-critical applications (science, climate studies etc.) the NetCDF formatted delayed-mode quality controlled data from the GDACs should continue to be used.

There are two approaches that could be taken: (i) add the additional bio-geochemical variables to BUFR Master Table 0 under Class 22 (for oceanographic variables), and to continue to use Master Table 0, or (ii) define a new template using Master Table 10. After discussions with BUFR experts within the Met Office, JCOMM and WMO it is clear that few (if any) operational centres have the ability to decode Master Table 10 encoded messages, such that in the near future Master Table 0 will have to be used.

BUFR Master Table 0 can accommodate additional near surface measurements (e.g. un-pumped temperatures or auxiliary measurements from STS floats) and dissolved oxygen profiles, as discussed below. For more exotic bio-geochemical variables a decision will need to be made whether to include the additional variables under Master Table 0 (Class 22) or to move to Master Table 10. However, for the time being the approach should be to continue to generate Argo BUFR messages based on Master Table 0.

Argo data system approach

The approach agreed by the Argo Data Management Team (ADMT) is for the Argo DACs (Data Assembly Centres) to create separate NetCDF files for the core (pressure, temperature, salinity, conductivity) parameters (core-files) and for the bio-geochemical data (bio-files); where a bio-file contains all parameters from a float, except for the core parameters (temperature, salinity, conductivity). Hence, a standard float that performs only CTD measurements will not have associated bio-files. This is detailed at http://www.argodatamgt.org/Documentation/Draft-documents.

However, some CTD-only floats can also measure specialized profiles with different vertical sampling schemes, such as:

- bouncing profiles: a series of shallow profiles performed during one cycle;
- high resolution near-surface observations: higher resolution vertical sampling near the surface with un-pumped temperature measurements.

For these the measurements would be stored as an additional profile (on a different vertical axis) within the NetCDF core-file.

Recommended approach for the Argo BUFR messages

It is recommended that a similar approach is adopted for the Argo BUFR messages, where supplementary (temperature/salinity) measurements and/or dissolved oxygen would be reported (in the same BUFR message) as additional data sequences, as discussed below and detailed in Appendix C.

a) high-resolution near surface sampling

For standard floats the CTD pump is normally switched off at around 5 dbar depth to avoid contamination of the salinity sensor. However, a number of float types continue to sample (with the pump off) up to the surface. This includes Apex (Argos) floats that report un-pumped temperature measurements (with nearly 400 such floats presently operating), Provor/Arvor/Nova floats that report both un-pumped temperature and salinity, and Apex STS floats that have a free-flushing auxiliary CTD sensor (with around 60 of these presently operating). For SOLO and S2A floats that keep the pump on closer to the surface, the existing template (3 15 003) can be used as normal.

i. high-resolution near surface temperature profiles

It is suggested that a new sequence is defined to allow for an additional higher resolution near-surface un-pumped temperature profile when these data are reported. The new sequence uses pressure as the vertical axis (consistent with the standard Argo template) as this is what is actually measured and includes a qualifier that allows the measurements to be identified as being from an un-pumped sensor. It is important to make this distinction as the error characteristics of the un-pumped temperatures will be different to those measured when the pump is on. The temperatures are only reported to 2 decimal places (rather than 3 decimal places as in the standard Argo temperature profile) as the un-pumped temperatures would be expected to be of slightly lower accuracy.

These measurements may overlap with the primary Argo CTD profile for the purpose of cross-calibration. However the values reported in the real-time NetCDF core-file and BUFR sequence would be unadjusted.

ii. high-resolution near surface temperature and salinity profiles

These can be reported from various floats (as noted above), either by sampling both temperature and salinity while the pump is off, or through an auxiliary free-flushed CTD sensor. However the salinity from STS floats would be expected to drift from fouling as it samples through the ocean surface, but overlapping data acquired from the primary CTD sensor should allow the STS salinities to be adjusted, profile by profile, to match the primary salinity measurements. However, as for the near surface un-pumped temperatures the auxiliary salinities reported from STS floats would be unadjusted in the real-time NetCDF core-file and BUFR sequence.

b) dissolved oxygen

A similar approach is suggested to allow the inclusion of dissolved oxygen profiles, by using an additional sequence (3 06 037) when these data are reported. This sequence has already been submitted to WMO and approved for validation. The sequence allows for both pressure and depth to be given as the vertical axis; it is suggested for Argo use that pressure is used (consistent with the standard Argo profile) and the depth elements are coded as missing.

For Argo data processing the dissolved oxygen sequence would be generated by the DACs using the oxygen data from the real-time NetCDF bio-file. A complication here is

that in the NetCDF core file the pressure has been adjusted (to account for any offset) while in the NetCDF bio-file the pressure is unadjusted (raw). However, the pressure variable will have exactly the same dimensions in both files, such that the pressure should map between the core and bio files, so the adjusted pressure should readily be obtained from the primary CTD profile in the core file. The adjusted pressure should be used as the pressure offset value is not reported in the BUFR sequence.

When used, the required sequences would be included in Section 3 of the BUFR message after the 3 15 003 and the additional data would follow the standard Argo CTD data in Section 4.

The advantage of this approach is that no changes would be needed when only the standard (pumped) temperature and salinity data are reported, template 3 15 003 can continue to be used as is. The additional sequences (one of which has already been accepted for validation) would then permit (any or all of) the additional near-surface un-pumped temperature (and salinity) measurements, auxiliary temperature and salinity measurements from STS floats and dissolved oxygen measurements to be included in the Argo BUFR message(s) disseminated on the GTS.

Appendix A

Format of the Argo BUFR message

An observation report coded as BUFR is a continuous stream of binary data. A BUFR message consists of 6 sections including the start (Indicator) and End sections which delimit the 4 inner sections. The next 2 sections (Section 2 is optional and reserved for local use, so is usually omitted) are for handling the BUFR message as a whole during transmission and give a rough classification of the data and a single "representative" time (this does not mean that the time can be omitted from the data itself). The Argo-specific data are contained in Sections 3 and 4, the data descriptor(s) and data values respectively, as described below.

SECTION 0	Indicator Section
SECTION 1	Identification Section
SECTION 2	(Optional Section)
SECTION 3	Data Description Section
SECTION 4	Data Section
SECTION 5	End Section

SECTION 3

This section defines the data that is to follow in section 4.

The first 3 octets (1 octet = 8 bits) give the length of the section (in octets). Octet 4 is reserved and set to 0 (00000000). Octets 5 and 6 represent the number of data subsets, this is normally 1 (00000000 00000001) as each float profile is generally sent individually.

Octet 7 defines whether this is observed data or other data. This should be set to 10000000 to indicate observed data (that is not compressed).

The data descriptors start in octet 8. Each data value has a descriptor based on an F-X-Y structure where each descriptor is 16 bits-wide. F refers to the two most significant bits (leftmost), X refers to the 6 middle bits and Y to the least significant (rightmost) 8 bits. Hence F has values in the range 0 to 3, X from 0 to 63 and Y from 0 to 255. Conventionally, when F = 0 then it refers to an element, F=1 refers to replication and F=3 to a template or a sequence (of descriptors).

For Argo data a single template descriptor **3 15 003** is presently used which specifies the form of the data that follows in Section 4. The template in turn describes the various sequences and descriptors that define the data, as given in Appendix B.

Hence this section is just 9 octets in length:

- 7 octets before the template descriptor (as described above)
- 2 octets for the template descriptor (11001111 00000011)

SECTION 4

This section contains the data as defined by order of the descriptors defined by the template. Again the first 3 octets give the length of the section; this will depend on the number of profile levels being reported from the float. Octet 4 is reserved and set to 00000000. The data starts in octet 5 in the order defined by the template. All the data is given as a continuous binary sequence. Each data point occupies a specified number of bits as defined in the template. Missing values are denoted by filling each data bit with 1s, e.g. for a temperature only float all salinity bits should be filled with 1s. The data should be padded with 1s at the end of the section to ensure the length of the section is an integer number of octets.

Appendix B

BUFR template 3 15 003 for temperature and salinity profiles from sub-surface profiling floats

0 01 087		WMO marine observing platform (extended 7 digit) identifier
0 01 085		Observing platform manufacturers model
0 01 086		Observing platform manufacturers serial number
0 02 036		Buoy type
0 02 148		Data collection and/or location system
0 02 149		Type of data buoy
0 22 055		Float cycle number
0 22 056		Direction of profile
0 22 067		Instrument type for water temperature profile measurement
3 01 011	0 04 001	Year
	0 04 002	Month
	0 04 003	Day
3 01 012	0 04 004	Hour
	0 04 005	Minute
3 01 021	0 05 001	Latitude (high accuracy)
	0 06 001	Longitude (high accuracy)
0 08 080		Qualifier for quality class
0 33 050		Global GTSPP quality class
1 09 000		Delayed replication of 9 descriptors (after delayed replication factor)
0 31 002		Extended delayed descriptor replication factor (gives # of vertical levels)
0 07 065		Water pressure (Pa)
0 08 080		Qualifier for GTSPP quality class
0 33 050		Global GTSPP quality class
0 22 045		Sub-surface sea temperature (K)
0 08 080		Qualifier for GTSPP quality class
0 33 050		GTSPP quality class
0 22 064		Salinity ($^{o}/_{oo}$)
0 08 080		Qualifier for GTSPP quality class
0 33 050		GTSPP quality class

The template defines the data in terms of the various descriptors, where each descriptor is associated with a scaling factor, offset and data width to enable a value to be reported using a specified number of data bits. These are documented in BUFR Table B, where the descriptors used in the Argo template are given on the following page.

Where the value is given by a (CCITT IA5) text string then each ASCII character is given by its associated decimal (range 0 to 255) which is represented by 1 byte (8 bits). Where the text string is shorter than the number of characters specified it is right-most padded with blank characters.

For a numerical value the number represented may be scaled and/or offset to minimize the number of bits needed to represent its range. For example descriptor 0 01 087 gives the 7-figure WMO number, this is represented with 23 bits.

The use of scaling can be illustrated by the coding of latitude, say the floats position is 51.6083N, then for latitude (0 05 001) we have sf = 5 and offset = -9000000 and a 25 bit data width, so for 51.6083N this codes to

$(51.6083 \times 10^5) + 9000000 = 14160830$,

which is represented in 25 bits.

Note the end of the section 4 data must be padded with 1s to ensure its length is given by an integer number of octets.

Descriptor	Data	Scaling factor	Offset	Data width
0 01 087	Numeric	0	0	23
0 01 085	20 characters			160
0 01 086	32 character			256
0 02 036	Numeric (see note 1)	0	0	2
0 02 148	Numeric (see note 2)	0	0	5
0 02 149	Numeric (see note 3)	0	0	6
0 22 055	Numeric	0	0	10
0 22 056	Numeric (see note 4)	0	0	2
0 22 067	Numeric (see note 5)	0	0	10
0 04 001	Numeric (year)	0	0	12
0 04 002	Numeric (month)	0	0	4
0 04 003	Numeric (day)	0	0	6
0 04 004	Numeric (hour)	0	0	5
0 04 005	Numeric (minute)	0	0	6
0 05 001	Numeric	5	-9000000	25
0 06 001	Numeric	5	-18000000	26
0 08 080	Numeric (see note 6)	0	0	6
0 33 050	Numeric (see note 7)	0	0	4
1 09 000	Indicates 9 descriptors (0 07 065 – 0 33 050) are repeated for each profile level			
0 31 002	Numeric (# of profile levels)	0	0	16
0 07 065	Numeric (Pa)	-3	0	17
0 08 080	Numeric (see note 6)	0	0	6
0 33 050	Numeric (see note 7)	0	0	4
0 22 045	Numeric (K)	3	0	19
0 08 080	Numeric (see note 6)	0	0	6
0 33 050	Numeric (see note 7)	0	0	4
0 22 064	Numeric ($^o/_{oo}$)	3	0	17
0 08 080	Numeric (see note 6)	0	0	6
0 33 050	Numeric (see note 7)	0	0	4

Note 1. Buoy type

Use value of 2 for sub-surface float.

Note 2. Data collection and/or location system

1 Argos
5 ORBCOMM
7 Iridium
8 Iridium and GPS
9 Argos-3
10 Argos-4

Note 3. Type of data buoy (note only those entries relevant to floats are given here)

8 Unspecified sub-surface float
10 ALACE
11 MARVOR
13 PROVOR
14 SOLO
15 APEX
26 Sub-surface Argo float
27 PALACE
28 NEMO
29 NINJA
30 Ice buoy/float (POPS or ITP)

Note 4.

0 Upwards profile
1 Downwards profile
2 Horizontal
3 Missing value

Note 5. Instrument type for water temperature profile measurement

831 CTD-P-ALACE float
837 ARVOR_C, SBE conductivity sensor
838 ARVOR_D, SBE conductivity sensor
839 PROVOR-II, SBE conductivity sensor
840 PROVOR, no conductivity sensor
841 PROVOR, Sea-Bird conductivity sensor
842 PROVOR, FSI conductivity sensor
843 Polar Ocean Profiling System (POPS), PROVOR, SBE CTD
844 Profiling float, ARVOR, Sea-Bird conductivity sensor
845 Webb Research, no conductivity sensor
846 Webb Research, Sea-Bird conductivity sensor
847 Webb Research, FSI conductivity sensor
848 APEX-EM, SBE conductivity sensor
849 APEX_D, SBE conductivity sensor
850 SOLO, no conductivity sensor
851 SOLO, Sea-Bird conductivity sensor
852 SOLO, FSI conductivity sensor
853 Profiling float, SOLO2 (SCRIPPS), Sea-Bird conductivity sensor
854 S2A, SBE conductivity sensor
855 Profiling float, NINJA, no conductivity sensor
856 Profiling float, NINJA, SBE conductivity sensor
857 Profiling float, NINJA, FSI conductivity sensor
858 Profiling float, NINJA, TSK conductivity sensor
859 Profiling float, NEMO, no conductivity sensor

860 Profiling float, NEMO, SBE conductivity sensor
861 Profiling float, NEMO, FSI conductivity sensor
862 SOLO_D, SBE conductivity sensor
863 NAVIS-A, SBE conductivity sensor
864 NINJA_D, SBE conductivity sensor
865 NOVA, SBE conductivity sensor
901 Ice-tethered Profiler (ITP), SBE CTD

Note 6. Qualifier for GTSPP quality class

0 Total water pressure profile
1 Total water temperature profile
2 Total water salinity profile
3 Total water conductivity profile
4 Total water depth
5-9 Reserved
10 Water pressure at a level
11 Water temperature at a level
12 Salinity at a level
13 Water depth at a level
14-19 Reserved
20 Position
21-62 Reserved
63 Missing value

Note 7. GTSPP quality class

0 Unqualified
1 Correct value (all checks passed)
2 Probably good but value inconsistent with statistics (differ from climatology)
3 Probably bad (spike, gradient, etc., if other tests passed)
4 Bad value, impossible value (out of scale, vertical instability, constant profile)
5 Value modified during quality control
8 Interpolated value
9 Good for operational use; caution for other use (check literature)
10-14 Reserved
15 Missing value

Appendix C

Additional sequences

3 06 YYY
Near surface depth (pressure) and temperature profile from floats

Table Reference			Table References			Element Name	
F	X	Y					
3	06	YYY				Sub-surface temperature profile (high accuracy / precision) with quality flags	
			0	02	032	Indicator for digitization	Set to 0, fixed sensor depths
			0	08	YYY	Temp/salinity measurement qualifier	
			1	06	000	Delayed replication of 6 descriptors	
			0	31	001	Delayed descriptor replication factor	Number of depths
			0	07	065	Water pressure	In Pa
			0	08	080	Qualifier for GTSPP quality class	
			0	33	050	Global GTSPP quality class	
			0	22	043	Sea / water temperature	In K to 2 decimal places
			0	08	080	Qualifier for GTSPP quality class	
			0	33	050	Global GTSPP quality class	
			0	08	YYY	Temp/salinity measurement qualifier	Set to missing to cancel

Code table 0 08 YYY is a new qualifier (4 bit width) for temperature/salinity measurement significance

where

1	Un-pumped float temperature and salinity data
2	Auxilliary STS sensor data
3 – 14	Reserved
15	Missing Value

3 06 ZZZ
Near surface depth (pressure), temperature and salinity profile from floats

Table Reference			Table References			Element Name	
F	X	Y					
3	06	ZZZ				Sub-surface temperature profile (high accuracy / precision) with quality flags	
			0	02	032	Indicator for digitization	Set to 0, fixed sensor depths
			0	08	YYY	Temp/salinity measurement qualifier	
			1	06	000	Delayed replication of 6 descriptors	
			0	31	001	Delayed descriptor replication factor	Gives number of depths
			0	07	065	Water pressure	In Pa
			0	08	080	Qualifier for GTSPP quality class	
			0	33	050	Global GTSPP quality class	
			0	22	043	Sea / water temperature	In K to 2 decimal places
			0	08	080	Qualifier for GTSPP quality class	
			0	33	050	Global GTSPP quality class	
			0	22	064	Salinity	
			0	08	080	Qualifier for GTSPP quality class	
			0	33	050	Global GTSPP quality class	
			0	08	YYY	Temp/salinity measurement qualifier	

3 06 037
Sequence for dissolved oxygen profile data

Table Reference			Table References			Element Name	
F	X	Y					
3	06	037				Dissolved oxygen profile data	
			1	09	000	Delayed replication of 9 descriptors	
			0	31	002	Extended delayed descriptor replication factor	Gives number of depths
			0	07	062	Depth below sea / water surface	Code as missing
			0	08	080	Qualifier for quality class	Code as missing
			0	33	050	GTSPP quality class	Code as missing
			0	07	065	Water pressure	
			0	08	080	Qualifier for quality class (set to 10, indicates pressure at a level)	
			0	33	050	GTSPP quality class	
			0	22	188	Dissolved oxygen	New entry - in µmol kg^{-1}
			0	08	080	Qualifier for quality class (set to 16, dissolved oxygen at a level)	
			0	33	050	GTSPP quality class	

Code depth related descriptors as missing as water pressure is used as the vertical axis.