|  |  |  |
| --- | --- | --- |
| WORLD METEOROLOGICAL ORGANIZATION  COMMISSION FOR BASIC SYSTEMS  -----------------------------  FOURTH MEETING OF  INTER-PROGRAMME EXPERT TEAM ON DATA REPRESENTATION MAINTENANCE AND MONITORING  GENEVA, SWITZERLAND, 30 MAY - 3 JUNE 2016 |  | IPET-DRMM-IV / Doc. 3.2 (13)  (30. 5. 2016)  -------------------------  ITEM 3.2  ENGLISH ONLY |

3. BUFR AND CREX

**Reporting of meaurement uncertainties in BUFR**

*Submitted by Alexander Kats (Russia)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Summary and Purpose of Document**

This document contains a discussion about a need in unified approach for reporting of measurement uncertainties in BUFR and ways for its implementation

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ACTION PROPOSED**

The meeting is invited to discuss the content of this document, accept the need in unified approach for reporting of measurement uncertainties in BUFR and advice on the way for its development

**DISCUSSIONS**

1. Background:

This consideration is inspired by the request came from GRUAN community. GRUAN is GCOS Reference Upper Air Network aimed for providing long-term reference observations of upper air essential climate variables. One of an attribute of reference observations is reporting measurement uncertainties along with each measurand. In respect to vertical upper-air profiles it means that uncertainty estimates should be vertically resolved. As not only climate community but NWP as well may benefit from using uncertainty estimates in their data assimilation schemes GRUAN was requested to report their observations (so called GRUAN data product) in near real time and the only anticipated way to accomplish this, naturally, BUFR.

At this moment under consideration is upper-air radiosonde GRUAN data product. Two possibilities are discussed: report compliant with B/C GRUAN radiosonde data product instead or along with (in separate bulletin) conventional manufacturer’s data processing or use dedicated GRUAN BUFR template (to be developed). Both option imposes result in challenge in respect to representation of uncertainty estimates.

Along with recently proposed templates for reporting results of radar, RASS and LIDAR/ceilometer vertical profiling, this request highlights a necessity to develop a universal generic approach for reporting measurement uncertainties in BUFR (and CREX?).

1. Precedents and their shortcomings

By the moment there are two dedicated descriptors in Table B Class 11 “Wind and turbulence” and Class 12 “Temperature” related to uncertainty.

They seem to be intended for use in 3 12 030 sequence for reporting satellite scatterometer winds:

0 11 052 Formal uncertainty in wind speed [m s–1]

0 11 053 Formal uncertainty in wind direction [degree true]

According to explanations [found](http://projects.knmi.nl/scatterometer/publications/pdf/RapidScat_Product_Manual.pdf) in KNMI documentation “The ‘Formal Uncertainty in Wind Direction’ does not contain the uncertainty, but the normalised inversion residual” and 0 11 053, as most likely as 0 11 052, seems to be useless for expressing measurement uncertainties.

The second group are recently proposed descriptors in the Classes 11, 12 and 15 which are dedicated namely to report “Measurement Uncertainty Estimate” as referred to by CIMO Guide it. I.1.6.2 “Non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used” in respect to concrete quantities, like:

0 11 110 Uncertainty in u-component [m s–1]

0 11 111 Uncertainty in v-component [m s–1]

0 11 112 Uncertainty in w-component [m s–1]

0 12 008 Uncertainty in virtual temperature [K]

0 15 064 Uncertainty in Attenuated Backscatter [m-1·Sr-1]

0 15 066 Uncertainty in Particle Backscatter Coefficient [m-1·Sr-1]

0 15 068 Uncertainty in Particle Extinction Coefficient [m-1]

0 15 070 Uncertainty in LIDAR Ratio [Sr]

0 15 072 Uncertainty in Depolarization Ratio [%]

When the Team considered 3 09 021 and 3 09 022 templates their proposers discussed a generic universal descriptor “Measurement Uncertainty Estimate” which was rejected due to associated problems such as how to define unit, reference, scale and width to make it appropriate for any arbitrary quantity.

However, allocation dedicated descriptors to each quantity to be reported potentially may result in fast exhausting of space available for allocation of new descriptors as it seems reporting uncertainties becomes a trend.

Another potential shortcoming is that uncertainty may be expressed in several ways, not necessarily, as assumed by default[[1]](#footnote-1), as standard uncertainty (standard deviation), but as well using interval estimates (e.g. expanded uncertainty) that may be important for non-Gaussian errors distribution, especially asymmetric. As well, because GRUAN specific requirement is detailed characterisation of uncertainties, component of total uncertainties may need to be reported such as systematic and random components (in relation to upper-air observations – vertically correlated and uncorrelated constituents).

1. Analogues

There are also several dimensional descriptors in Table B for reporting errors, specific to observed quantities, such as

0 15 032 Estimated error in atmospheric path delay [m]

0 15 034 Estimated error in path delay difference [m]

0 40 002 Estimated error in surface soil moisture [%]

and being validated

0 40 031 Error estimate of horizontal line of sight wind [m s–1]

They all have the same limitations as discussed above.

An alternative approach for characterisation of errors is demonstrated in the sequence 3 10 067 “Satellite-derived winds”, which is under validation where significance qualifiers from Class 8 Code table 0 08 041 “Data significance” precede descriptors, describing data elements, to which error estimates are applied:

0 08 041 Data significance (14 = Expected error)

0 11 002 Wind speed

0 08 041 Data significance (15 = Representative error)

0 10 004 Pressure

0 12 101 Temperature/air temperature

0 08 041 Data significance (Missing = Cancel)

In the Class 8 there is also a similar in some way Code Table 0 08 023 “First-order statistics” although it applies to a measurands itself rather than to its errors, e.g.:

Code figure

…

4 Mean value

7 Mean absolute error

…

9 Best estimate of standard deviation (N–1)

10 Standard deviation (N)

…

12 Root-mean-square vector error

13 Root-mean-square

…

Such approach could be applied for reporting uncertainties after defining a proper data significance descriptors in the Class 8 with the same descriptors used for reporting respective measurands thus resolving issues with defining unit, reference, scale and width.

Among those built-in BUFR mechanisms, which usage of Table C operators provides for reporting quality assessment information, there are two ones that serve for tasks similar to reporting of measurement uncertainties.

The first one is the adding associated fields with Table C operator 2 04 Y followed by descriptor 0 31 021 “Associated field significance” from the Class 31 “Data description operator qualifiers”. So far most of entries in Code table 0 31 021 are for qualitative quality control information, e.g.:

…

1 1-bit indicator of quality 0 = good

1 = suspect or bad

2 2-bit indicator of quality 0 = good

1 = slightly suspect

2 = highly suspect

3 = bad

…

But in the Code table 0 31 021 there is also a quantitative entry 7 “Percentage confidence” which allow preceding reported data values with numeric estimates of their confidence that in some way similar to the idea of reporting uncertainties.

Another Table C operator 2 22 000 “Quality information follows” allows to report quality information with the values of elements of the Class 33 “Quality information” addressing preceding data elements, to which the information applies, with data present bit-map. This provides efficient way to report quality information for data conveyed using Table D sequences. So far, the Class 33 contains mostly qualitative descriptors defined by respective Code and Flag tables except several ones, e.g.:

0 33 007 Per cent confidence [%]

…

0 33 009 Relative error [%]

…

0 33 040 Confidence interval [%]

0 33 089 Noise equivalent delta temperature (NEdT) quality indicators for warm target calibration [K]

0 33 090 NEdT quality indicators for cold target calibration [K]

0 33 091 NEdT quality indicators for overall calibration [K]

Especially of interest “generic” descriptors 0 33 009 and 0 33 040 as they seem to provide a way for reporting dimensionless quantitative value characterising performance of a value of arbitrary dimension. However there are neither explanations about application of these descriptors nor examples of its usage in any templates in use or under validation were found.

1. Speculations
   1. Using significance qualifiers

The most natural way for reporting measurement uncertainty seems to be introducing respective significance qualifiers into Class 8 and apply them to the same data descriptors used for reporting observables (followed by respective cancellation), when the meaning of a data element is clear from a context. Two Class 8 descriptors defined by code tables may be required such as (preliminary):

0 08 092

*Measurement uncertainty expression*

Code figure

…

1 Standard uncertainty

2 Expanded uncertainty at 95% confidence level

3 Expanded uncertainty (K=2)

4 Expanded uncertainty (K=3)

5 Coverage interval at 95% confidence level

6 Lower bound of coverage interval at 95% confidence level

7 Upper bound of coverage interval at 95% confidence level

…

All bits set to zero Missing value

0 08 093

*Measurement uncertainty significance*

Code figure

…

1 Total uncertainty

2 Systematic component of uncertainty

3 Random component of uncertainty

…

All bits set to zero Missing value

Such an approach solves all issued with defining unit, reference, scale and width. When required, reference, scale and width may be redefined using respective Table C operators.

* 1. Using associated fields

When the meaning of a data element may be unclear from a context it may be considered using associated fields for which is necessary to introduce respective entry to Code table 0 31 021

9 Measurement uncertainty

with a Note (4) “The code figure 9 specify expression of measurement uncertainty using the same unit, reference, scale and width used by a data element with which the field is associated. The meaning of the field shall be specified with significance qualifiers 0 08 092 and 0 08 093 to be implicitly[[2]](#footnote-2) cancelled immediately after respective associated field”.

An example of respective representation of air temperature reported with 0 12 101 is given below:

2 04 016+N Add 16 bit+N associated field, where N= n(0 08 092)+n(0 08 093)[[3]](#footnote-3)

0 31 021 = 9 measurement uncertainty

0 08 092 = 1 standard uncertainty

0 08 093 = 1 total uncertainty

0 12 101 = temperature reported

2 04 000 = cancel add associated field

This sequence assumes reporting in Data section 3 6+n(0 08 092)+n(0 08 093)+16+16 bits where the first occasion of 16 bits is to be used for reporting temperature measurement uncertainty with the second one for reporting temperature itself. However, it’s unclear how legal is using of 0 08 092 and 0 08 093 in such a context. End, even if it’s legal - it requires modification of existing decoders. Nonetheless, defining e.g. 0 31 021=9 as “standard total measurement uncertainty to be expressed with the same unit, reference, scale and width used by a data element with which the field is associated” shall work.

Such approach may present even more difficulties when several components of uncertainties need to be reported. It may be possible to achieve it using nesting of 2 04 Y operators but it seems to be too complicated.

* 1. Dealing with existing templates

The next issue is a reporting measurement uncertainties for values to be reported using already existing Table D sequences not intended for their conveying. It seems there are two ways to achieve this without inventing new sequences.

* + 1. Reporting uncertainty repeating coordinate descriptors

The very first one is to reproduce the descriptors for data elements to be characterized by measurement uncertainties with preceding significance qualifiers 0 08 092 and 0 08 093. To resolve several data elements conveyed with the same data descriptors within the same Table D sequence (e.g. like multiple repeating 0 12 101 temperature in TM309052) it will be necessary to reproduce as well respective coordinate descriptor (or descriptors if necessary from context), for example:

1 01 000 Delayed replication of 1 descriptor

0 31 002 Extended delayed descriptor replication factor

0 04 086 Long time period or displacement

0 08 092 = 1 standard uncertainty

0 08 093 = 2 systematic uncertainty

0 12 101 systematic component of temperature standard uncertainty

0 12 103 systematic component of dewpoint temperature standard uncertainty

0 08 093 = 3 random uncertainty

0 12 101 random component of temperature standard uncertainty

0 12 103 random component of dewpoint temperature standard uncertainty

0 08 093 = all bits set to zero

0 08 092 = all bits set to zero

* + 1. Using data present bitmap

Another idea is to define new Class 33 descriptors for use with Table C operator 2 22 000 “Quality information follows”.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TABLE |  |  |  |  | DATA |
| REFERENCE | ELEMENT NAME | UNIT | SCALE | REFERENCE | WIDTH |
| F X Y |  |  |  | VALUE | (Bits) |
| 0 33 092 | Relative measurement uncertainty (see Note 4) | % | 2 | 0 | 12 |

Notes:

(4) When using descriptor 0 33 092, the range to which its value shall apply shall be determined by the range of respective data element representation addressed by preceding bit-map.

The advantage of using data present bitmap is that it directly relates uncertainty estimates with observed values instead of necessity to repeat coordinate descriptors (e.g. time as in example above) to make a link.

Expressing relative uncertainty in per cent allows universal representation of uncertainty regardless of parameters defining BUFR representation of respective values (scale, reference and width). To avoid singularity (e.g. when reporting relative uncertainty for a data value equal to zero) divisor shall to be defined as non-zero value. The most natural solution for achieving this looks using range of definition of reported value. Shortcoming that naturally uncertainty is much less of the measurement range. For example, for reporting pressure with 0 07 004 descriptor 14 bits are used, scale is -1, i.e. total range is 16384 \* 10 Pa or 1638.4 hPa. Thus, 1 hPa absolute uncertainty typical for upper-air observations nearby surface gives 0.06 % relative uncertainty while 0.04 hPa absolute uncertainty typical for upper-air observations at 10 hPa gives (roughly) 0.02%. So, it’s clear that using Table C operators for modifying width, scale and even reference value (to signify lower bound or both bounds of coverage interval when it or they are negative) for 0 33 092 in some cases will be required.

An example of sequence for reporting systematic (vertically correlated) and random components of uncertainty of temperature and dewpoint temperature after conveying sequence 3 09 052 may look like:

2 22 000 Quality information follows

0 08 092 = 1 standard uncertainty

0 08 093 = 2 systematic uncertainty

2 36 000 Define data present bit-map

1 01 000 Delayed replication of 1 descriptor

0 31 002 Extended delayed descriptor replication factor = amount of data values to be addressed by the data present bit-map

0 31 031 Data present indicator (bits one by one creating bit-map where zeroes indicate referred data values)

0 33 092 Relative measurement uncertainty [%]

2 37 000 Use defined data present bit-map

0 08 093 = 3 random uncertainty

0 33 092 Relative measurement uncertainty [%]

0 08 093 = all bits set to zero

0 08 092 = all bits set to zero

2 35 000 Cancel backward data reference

Here a bitmap constructed by repeating 0 31 031 should represent a pattern like:

Npl\*1111110011+1+Nsl\*1111111

where Npl is amount of data levels (i.e. amount of repetition of 3 03 054) and Nsl is amount of levels with wind shear data (i.e. amount of repetition of 3 03 051). Last but not least – with 1 sec resolution addressing 65535 data elements possible with 0 31 002 may be insufficient.

An alternative to using relative uncertainty is to put to class 33 all required quantities:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TABLE |  |  |  |  | DATA |
| REFERENCE | ELEMENT NAME | UNIT | SCALE | REFERENCE | WIDTH |
| F X Y |  |  |  | VALUE | (Bits) |
| 0 33 xxx | Pressure measurement uncertainty | Pa | 2 | 0 | TBD |
| 0 33 xxx | Temperature measurement uncertainty | K | 2 | 0 | TBD |
| 0 33 xxx | Relative humidity measurement uncertainty | % | 1 | 0 | TBD |
|  | etc |  |  |  |  |

One more option is to introduce new Table C operator 2 XX 000 “Uncertainty information follows”, but this again requires upgrade of decoders.

1. Further issues

The consideration above was limited to scalar values. Most common example of vector quantity is wind. Reporting uncertainty for wind direction and wind speed isn’t natural although possible. Wind uncertainty is better characterized by standard vector wind deviation (root-mean square vector wind deviation) in case of uniform spatial distribution errors but may require more sophisticated approaches (in former times it was found that differences between radar and navaid OMEGA wind form ellipse with orientation demonstrating some diurnal variation due to different conditions for OMEGA signals propagations).

There are already more sophisticated cases covered in existing Table B such as lighting location errors characterized by following descriptors.

0 20 111 x-axis error ellipse major component

0 20 112 y-axis error ellipse minor component

0 20 113 z-axis error ellipse component

0 20 114 Angle of x-axis in error ellipse

0 20 115 Angle of z-axis in error ellipse

This approach may be actual as well for 3D wind components uncertainty reporting along with wind profilers data.

**PROPOSAL**

At the moment ideas on reporting uncertainties are not mature enough and requires further discussion both within the Team and with CIMO and, possibly, BIPM.

Therefore the meeting is invited to accept the need in unified approach for reporting of measurement uncertainties in BUFR (and, possibly – CREX), consider proposed ways for its development and make respective advices.

References:

1. Immler et al: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products. Atmospheric Measurement Techniques, 2010, 3, 1217–1231
2. Dirksen et al: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463-4490

1. There may be a need to state somewhere in WMO regulatory documents (or event directly in FM 94) that “uncertainty” by default means standard uncertainty. [↑](#footnote-ref-1)
2. To avoid violation of Note (7) under Table C [↑](#footnote-ref-2)
3. Here n(0 08 092) and n(0 08 093) refer to data width of respective code tables so far undefined [↑](#footnote-ref-3)