Progress report on migration to BUFR

by Bruce Ingleby and Drasko Vasiljevic, ECMWF, 13 March 2015

Executive Summary

The migration of surface and radiosonde observations to BUFR is a very large undertaking and progress has been very uneven, both by data producers and data users.

The main findings of this report are:

- In November 2014 alphanumeric SYNOP reports from the UK, Ireland and Netherlands ceased general circulation on the GTS as did alphanumeric TEMP SHIP reports from about half the ships making radiosonde ascents.
- Some countries are yet to produce any BUFR, overall about 30% of land and radiosonde stations are not generating BUFR
- There are fewer land surface reports in BUFR than TAC for some countries, mainly because only those stations/times nominated for international exchange are available in BUFR whereas extra reports are available in TAC. (One or two countries provide extra reports in BUFR.)
- For land surface stations most BUFR reports have been reformatted from SYNOP, this generally gives acceptable data, but there are some advantages to native BUFR.
- Reformatting of radiosonde TEMP reports is more problematic, a) there are significant numbers of reformatting errors and b) many reports are still in "parts", violating BUFR regulations and causing problems for data users – major NWP centres cannot make good use of such non-regulation BUFR data which is not what they were expecting. It is necessary to make some data producers aware that the reports are non-regulation.
- About 15% of radiosonde stations are producing native BUFR, often high resolution and mostly of good quality. The proportion is expected to increase significantly over the next year. Such reports offer advantages in precision, vertical resolution and in treatment of balloon drift.
- There are various other quality problems with both surface and radiosonde BUFR reports, including position errors, special cases in code tables and wrong units of wind speed
- New marine surface BUFR templates are not yet implemented.
- There can be minor temperature offsets when comparing TAC and BUFR reports which climate users should be aware of.

Most NWP centres have started using some land surface data in BUFR format, take up of BUFR radiosonde data is slower. Extensive work is required to check data quality before starting to use new BUFR subsets operationally – overlap of TAC reports and the latest version of BUFR reports is needed for rigorous quality checking. If something can go wrong in observation coding/decoding then it usually does in some subset of the reports, there are both recent and historical examples of this leading to poorer forecast skill. This leads to a cautious approach and extensive change processes for NWP systems (which can delay the take up of BUFR). A clear procedure (with two months' notice) for the withdrawal of TAC subsets would be very helpful. The WMO website about the migration should be kept up to date.

1. Introduction

Since 2002 a migration from traditional alphanumeric codes (TAC) to BUFR (a table driven binary code) has been underway coordinated by the World Meteorological Organization (WMO). This is the largest change to the reporting of *in situ* meteorological observations for at least 50 years and involves experts in observations, coding, databases and data assimilation at each National Meteorological Service. For surface reports the basic content of TAC and BUFR reports is much the same, with BUFR containing some extra metadata. For radiosonde reports there is a more fundamental change in order to move away from some of the restrictions and features of the older codes – features that were introduced when telecommunication speed and costs were overriding concerns. As well as higher precision and extra metadata the BUFR templates support much higher vertical resolution and the reporting of the position and time at each radiosonde level. There was a deadline of November 2014 for the cessation of TAC reports on the GTS (Global Telecommunications System) – this was not met although some small subsets of TAC reports have been switched off. As described in this report progress towards the dissemination of good quality BUFR reports is very uneven, with more problems for radiosonde reports than land surface reports. In early 2015 some Numerical Weather Prediction (NWP) centres are using subsets of the new BUFR reports and others are not. NWP centres have been exchanging information about data availability and quality in various ways: directly, via WMO and via an ECMWF web page.

The migration will be complete when BUFR (or CREX) reports have replaced TAC on the GTS and all users (notably NWP centres) are using the new format data which are at least as good, in terms of quality and quantity, as the old format data were. (For radiosonde data there has been some discussion of interim solutions – which would cause more work for users - so that most TEMP reports could be retired in the near future; by definition the migration would not be complete until the final solution is reached.)

Section 2 of this report provides a brief introduction to BUFR templates, coding and decoding issues. Sections 3, 4 and 5 examine land surface, ship and radiosonde reports, respectively. For each category there are issues of data availability (and timeliness), metadata and the quality of the report contents. Sections 6 and 7 discuss the use of BUFR data within NWP and the migration process (including the need for overlap and quality checking of TAC and BUFR reports). Section 8 provides a summary and look ahead. Appendices provide links to further information and details of some issues. This report attempts to be as comprehensive and up to date as possible, but because of the large number of different data producers and ongoing changes to the BUFR reports this is a very large task and for brevity some of the descriptions are only a summary of a complex situation.

2. BUFR templates, coding and decoding issues

In principle BUFR is only one of two Table Driven Code Formats – the other being a character version called CREX. In practice there seems to be little or no use of CREX.

There can be errors in BUFR decoders as well as in the reports and different decoders can vary in how they cope with unusual or non-standard BUFR reports. Ideally once a problem is suspected it should be confirmed with another decoder, perhaps at a different NWP centre.

A list of specific coding and decoding issues has been identified:

- BUFR radiosonde reports from Mexico use BUFR edition 3, rather than BUFR edition 4 as they should do.
- Some BUFR reports use an expanded set of descriptors rather than the Table D top level descriptor (for example BUFR surface reports from Japan, the table D entry was not defined when Japan started sending the reports).
- BUFR land surface reports use four different templates, all with fairly similar contents. Some countries are using two different templates (producing near-duplicates?): Czech Republic, Malaysia, Syria, Jordan, Saudi Arabia, Kuwait, Bahrain, Qatar, UAE, Oman and Yemen.
- In TAC there are about 150 stations making "MOBILE SYNOP" reports (these are used at the Met Office): they have character identifiers and report position each time. These are provided for in the BUFR B/C 5 document, but it is not known if these reports are currently available in BUFR.
- For ships and buoys new templates are (belatedly) under development or adoption. Currently BUFR ship bulletins still contains many BUOY reports which should move to a separate template. The new templates should also give a cleaner separation of moored and drifting buoy reports. It is highly desirable that SHIP and BUOY TAC reports should continue for now and overlap with the new templates: a) having old and new versions of the reports is very useful for detecting problems and b) otherwise we have two migrations in practice.
- Almost all BUFR radiosonde reports use template 309052 but the UK currently uses 309055 for high resolution reports. Because of two minor problems with the 309055 reports and the fact that most NWP users prefer to have pressure reported the UK is in the process of changing to template 309052. One advantage of 309055 is that it explicitly includes extra metadata (including radiosonde serial number and software version, potentially useful for reanalysis and climate users). Vaisala reports often append similar extra metadata to 309052 reports but unless action is taken by the data users this may be discarded by the decoding software.

2.1 Radiosonde reporting in TAC and BUFR

Some aspects of radiosonde reporting in TAC (TEMP/PILOT) are based around requirements in the mid-20th century (and are no longer so relevant):

- The need to keep telecommunication messages as short as possible; hence the use of a) 'significant' levels to summarise the entire profile, b) the reporting of temperature to a precision of 0.2°, with the tenths digit also used to indicate the sign of the temperature and c) the omission of the leading digit when reporting height.
- The importance of standard levels – for standard level analysis charts. Modern NWP systems can use data at any level, and relatively high resolution reporting is desirable. Standard levels still have a place in verification and observation based climate studies.

In TEMP/PILOT code the levels are split into standard and significant levels and reported between the surface and 100 hPa (parts A and B) and above 100 hPa (parts C and D). Wind-only (PILOT) reports generally had radar height as the vertical coordinate – this resulted in the need for heights corresponding on average to standard pressure levels to be defined (in the WMO Manual on Codes Volume II, Regional Codes and National Coding Practices; the heights used vary by region or country and sometimes by station and month). Significant level PILOT winds are reported with height as the vertical coordinate giving an unsatisfactory mixture of reported vertical coordinates between standard and significant levels (at the Met Office there is an attempt to reinsert the original height corresponding to the standard pressure levels). Partly because of the rise of GPS-radiosondes the use of radar is declining, and so is the number of wind-only ascents (but they are still common in a few countries). A further complication is that a few countries making full radiosonde ascents send the temperature and humidity information in TEMP and the wind information in separate PILOT reports.

When the BUFR regulations were drawn up (circa 2005) there was a desire to move towards high vertical resolution reporting and also to simplify the reporting structure. Hence it was decided that there should be a single report containing all data from the ascent, for timeliness this is preceded by a report when the radiosonde reaches 100 hPa (see BUFR regulations B/C20 and B/C25). For wind-only ascents the whole profile should be reported using either a height vertical coordinate (template 309051) or a pressure coordinate (template 309050). As an interim measure TEMP parts A, B, C and D from some countries are being reformatted into BUFR and transmitted on the GTS as separate parts (in violation of the BUFR regulations). There is no easy way to distinguish part A and part B from the BUFR header (because separate parts do not fit the BUFR schema) causing problems for data users. For example the Met Office databank duplicate report check is switched off for BUFR radiosonde reports – but disabling the duplicate check is not considered safe for operational implementation.

Many BUFR reports have been produced by reformatting (or converting) TAC reports, although the proportion generated directly from the raw data ("native BUFR") should rise over time. For SYNOPs the reformatting generally gives acceptable BUFR reports but for radiosondes it causes a number of problems (see sections 5 and 6).

3. Land surface (SYNOP) stations

3.1 Availability and metadata

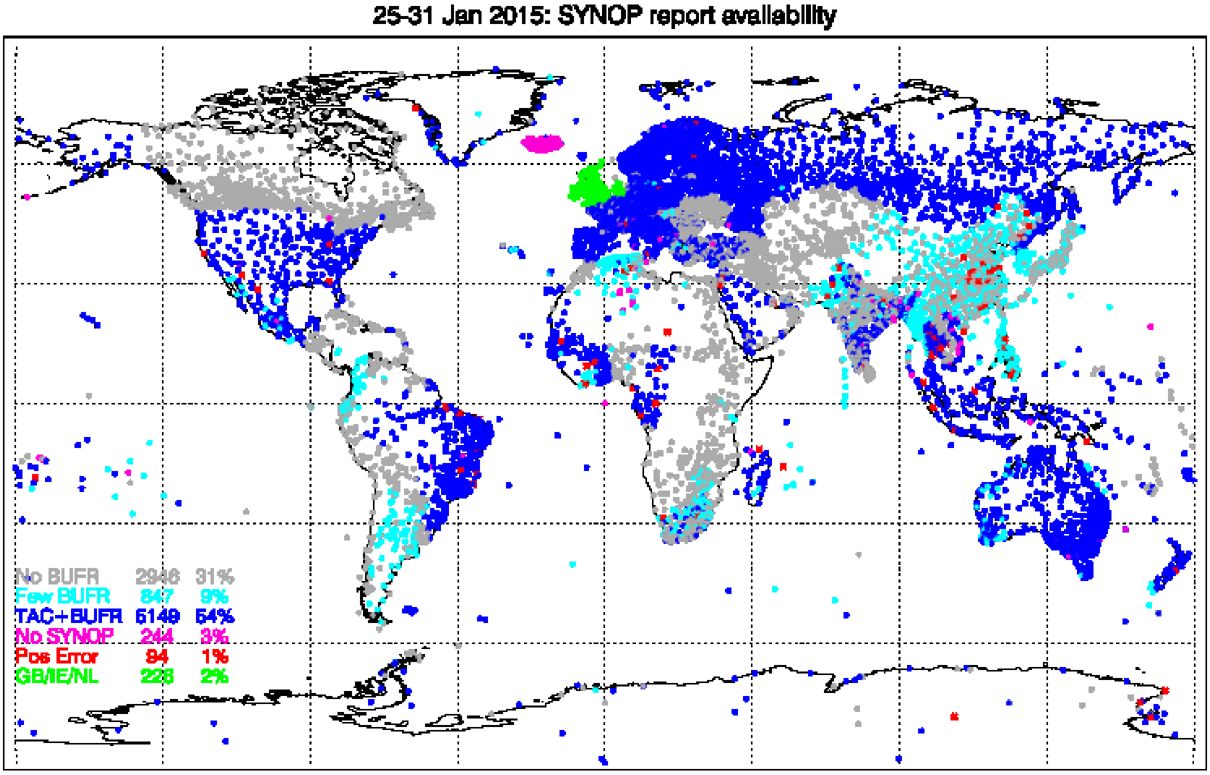

Figure 1. Availability of TAC SYNOP and BUFR land surface reports at ECMWF for 25-31 January 2015. Grey – no BUFR, blue - both TAC and BUFR (light/dark blue – less/more than 60% of TAC reports available in BUFR), purple – no TAC, green – reports from UK/Ireland/Netherlands (TAC SYNOPs ceased on GTS in November 2014). Red symbols indicate minor position errors (differences from WMO publication 9A) – large position errors have been corrected. (Note: There are many 'extra' Brazilian reports in a different BUFR template not shown in this figure.)

Figure 1 summarises the reporting of TAC and/or BUFR by surface land stations worldwide. Regions without BUFR reports include a swathe from Iran to Kazakhstan (taking in Turkmenistan, Uzbekistan, Kyrgyzstan, Tajikistan and Afghanistan), Ukraine, Sri Lanka, much of East Africa from Egypt to Mozambique; Botswana, Namibia, Angola, Morocco; Canada, much of the Caribbean; Guatemala, Honduras, Nicaragua, Costa Rica, Panama, Venezuela, Guyana, Surinam, Peru, Bolivia, Paraguay, Chile and various Pacific Islands. Very few BUFR land surface reports were received from Niger, Nigeria and Libya. (No surface reports in either format were received from some countries including Afghanistan, Somalia, the Democratic Republic of the Congo, South Sudan and Nicaragua.) In late January 2015 we were notified that Canada was about to start transmitting surface BUFR from some stations. China, Japan, India, Oman, South Africa, Romania and Turkey provide BUFR land surface reports from less than 80% of the stations reporting TAC SYNOP. In some cases the BUFR may contain Regional Basic Synoptic Network (RBSN) stations and the TAC may also include additional "national" stations.

Where there are fewer reports per station in BUFR than TAC (light blue dots in figure 1) the main reason is that the BUFR reports are 6 hourly, whereas the SYNOP reports are 3 hourly. This applies in China, South Korea, Myanmar, Pakistan, Maldives, Algeria, South Africa, Zimbabwe, Argentina, Colombia and Ecuador (for Myanmar 18 UTC reports are also missing from BUFR, for Pakistan and Argentina the BUFR reports are at 03, 09, 15, 21 UTC rather than the main hours). For the Philippines both TAC and BUFR are available 3 hourly but the BUFR data stops during 28 January 2015. SYNOP reports from Japan and French Polynesia are available hourly but the BUFR is 3 hourly. Some of these availability issues could be due to GTS routing issues, or a restriction to the RBSN.

Countries with "new" stations reporting in BUFR but not in SYNOP include Brazil, Iceland and to a lesser extent Australia and India.

In a few cases stations have missing (or 00000) WMO identifiers. Marine reports have always included position information but for land stations the latitude, longitude and station height have come from WMO Publication 9A. In BUFR these details should be included with every report (in practice they may be inserted by software at the collecting centre rather than at individual stations). Various problems have been seen:

- Missing positions
- Positions changing from one report to another (either position being retyped or possibly due to two PCs at a station with inconsistent data)
- Differences from station list (incorrect conversion from degrees and minutes to decimal degrees, errors in sign of longitude or latitude, mistyping, sometimes differences between surface and radiosonde positions)

Position errors can be very damaging for NWP, so some assimilation systems have taken action and reset positions to those in the station list, either for all stations or for those where the reported position is more than a specified tolerance from the station list position. Station height values are also important especially in the processing of station level pressure reports. It appears that most station height values coded in BUFR are correct but more checking of this point is required.

3.2 Report contents and quality

TAC reports of temperature are in degrees Celsius, with a precision of 0.1° for surface reports (and 0.2° for radiosonde reports). BUFR reports are in degrees Kelvin with a precision of 0.01°. When comparing the two there can be small systematic offsets due to a) the Celsius to Kelvin conversion used and b) any allowance for rounding of the values in the TAC reports. In BUFR dew point temperature is reported (rather than dew point depression as in TAC). In both TAC and BUFR a small proportion of surface stations report relative humidity rather than dew point. For wind speed there can be issues in a conversion from knots to m/s (and occasionally the conversion isn't performed giving an error of a factor close to 2).

In traditional manual reports the different measurements are made over the 10-15 minutes prior to the nominal observation time. In the BUFR regulations B/C 1: "B/C 1.2.2.1 If the actual time of observation differs by 10 minutes or less from the standard time reported in Section 1, the standard time may be reported instead of the actual time of observation" and "The actual time of observation shall be the time at which the barometer is read."

In most cases the **report times** of the SYNOP and BUFR match (reports on the hour). For the USA the BUFR times are 4 to 11 minutes before the hour, for Mexico they can be up to 20 minutes before the hour (larger time differences would not have been matched up). For the cases examined (in January 2015) the temperatures of these reports matched up. These relatively small changes in report times are not really a problem but might cause a particular report to be used in a different assimilation window. For Finland the reverse situation was found: SYNOP and BUFR reports at the same time could have slightly different temperatures. The explanation (from FMI) is that the SYNOP reports are made at 10 minutes to the hour, the BUFR reports are made on the hour. Similar small differences were seen at a few Austrian stations (apparently from manual and automated reports at slightly different times combined into a BUFR report).

There are some examples of erroneous **temperature** differences, particularly from India (and Seychelles). From the sample examined it seems that there are two BUFR versions of most Indian reports - a correct one (including pressure) and one without pressure but with the temperature and dew point 0.3° too high. The reports with the temperature offset were first noted by A Maycock (Met Office).

It seems that most BUFR reports have been reformatted from SYNOP (or possibly METAR) code. Reports containing wind directions other than a multiple of 10° (as in SYNOP code) were examined and were found to come mainly from Scandinavian countries and Brazil – these are presumably generated direct from raw data (but this doesn't prove that all other reports are reformatted). The reformatting process causes some problems but fewer than for radiosondes (as discussed below).

Wind speed differences were examined in detail for one day (26 January 2015). We would expect some differences (up to 0.5 m/s) because some countries report wind speeds in knots in TAC. In BUFR reports are all in m/s but to one decimal place (unfortunately many countries are not using the decimal place). There are some rounding differences a bit larger than would be expected (for example most French and Swiss TAC speeds seem to have been rounded up in 1 m/s intervals). In a few countries there seems to have been confusion between knots and m/s (and possibly km/hour), it can be difficult to tell which is right but the following are educated guesses: Croatia – some BUFR winds in knots, Madagascar – some BUFR winds in knots, Russian Antarctic stations – BUFR winds in knots (unwanted factor of 2 in reformatting?), South Korea – some TAC winds in knots but labelled as m/s; Kenya – two stations with BUFR speed 0.1 times TAC speed (software problem?).

For most stations **pressure** should be reported both at station level (Pstn) and at mean sea level (Pmsl). In most cases these compare well between TAC and BUFR but there are a few examples of missing leading digit (notably for Pmsl from Pakistani stations, eg 19.3 hPa in BUFR when it should be 1019.3 hPa). For Pstn there are modest (0.2 to 0.5 hPa) differences between TAC and BUFR for Estonian stations, related to differences in station height, but the details are unclear. There are a few isolated differences for individual reports (mainly from India, Pakistan and Africa) possibly related to corrected reports. There were differences of about 0.5 hPa in Pmsl for two Brazilian stations 81717 and 81909 *(are these actually separate stations – new AWS BUFR data?).*

Other variables such as visibility, cloud and radiation have not been examined.

When reporting of snow depth was introduced in SYNOP code it was for non-zero values only (Code Table 3889 forbids the use of sss=000). This gives an unfortunate ambiguity - there is no way to distinguish "no data" from "zero snow". Zero snow (on the ground) is a useful observation, "no data" is not. In the reformatting from SYNOP to BUFR some countries were converting missing snow depth values to zero values in BUFR - this can be quite misleading when there is snow on the ground and in general missing SYNOP snow depths are now converted to missing BUFR snow depths. In Code Table 3889 a value of 001 is 1 cm of snow, up to 996 - 996 cm. The table also has some special cases: 997 - Less than 0.5 cm; 998 - Snow cover, not continuous; 999 - Measurement impossible or inaccurate. We have seen these converted to values of almost 10 m of snow in the BUFR reports. Reports generated directly in BUFR should report zero snow depths without ambiguity and also without spurious ~10m snow depths. There are a few additional snow depth reports from China in BUFR compared to SYNOP reports. (Some countries use a separate "snow depth only" BUFR template, this is not affected by the migration.)

4. Ship reports

In current ship bulletins, both TAC and BUFR, there are both moored buoys (distinguished by a five digit callsign) and ships. In January 2015 there were 317 buoys reporting in both TAC and BUFR; 2111 ships reporting in TAC and 1930 in BUFR (counting distinct callsigns – some of the callsigns are spurious, caused by typing errors). Twenty frequently reporting Norwegian rigs/platforms (with callsigns LF??) were received in BUFR but not in TAC but overall there are more ship reports in TAC than BUFR. Some reports are duplicated, apparently because both Météo-France and NOAA are putting BUFR versions onto the GTS.

In BUFR position errors were noted for several ships: wrong sign of latitude in some cases when longitude = 0° (this came from a reformatting program in Washington, it may have been fixed now but this is still to be confirmed). In both TAC and BUFR a few ships (C6OB, C6RN3, C6YA5, ONGA, S6ES6 and WTEY) appeared to make distinct reports at the same time. A number of ships (especially with seven character callsigns starting with W, e.g., WXS6134, WXU3434 and WXY2616) reported pressure in BUFR but not in TAC (some examples also seen by Météo-France).

Jon Turton (Met Office, 13 Feb 2015) wrote "The BUFR messages currently being distributed (BC/10 3-08-009) are generated from the TAC (FM-13). This is a temporary solution until we have the systems in place to generate the newer templates directly from raw (or native) data and the TAC2BUFR 3-9-009 messages will be progressively replaced. Our ship AWS (AMOS) data should be the first to be replaced in the next few months." The UK ceased general dissemination of its TAC SHIP reports in early March 2015 but will make them available to other centres on request.

5. Radiosonde reports

5.1 Availability and metadata

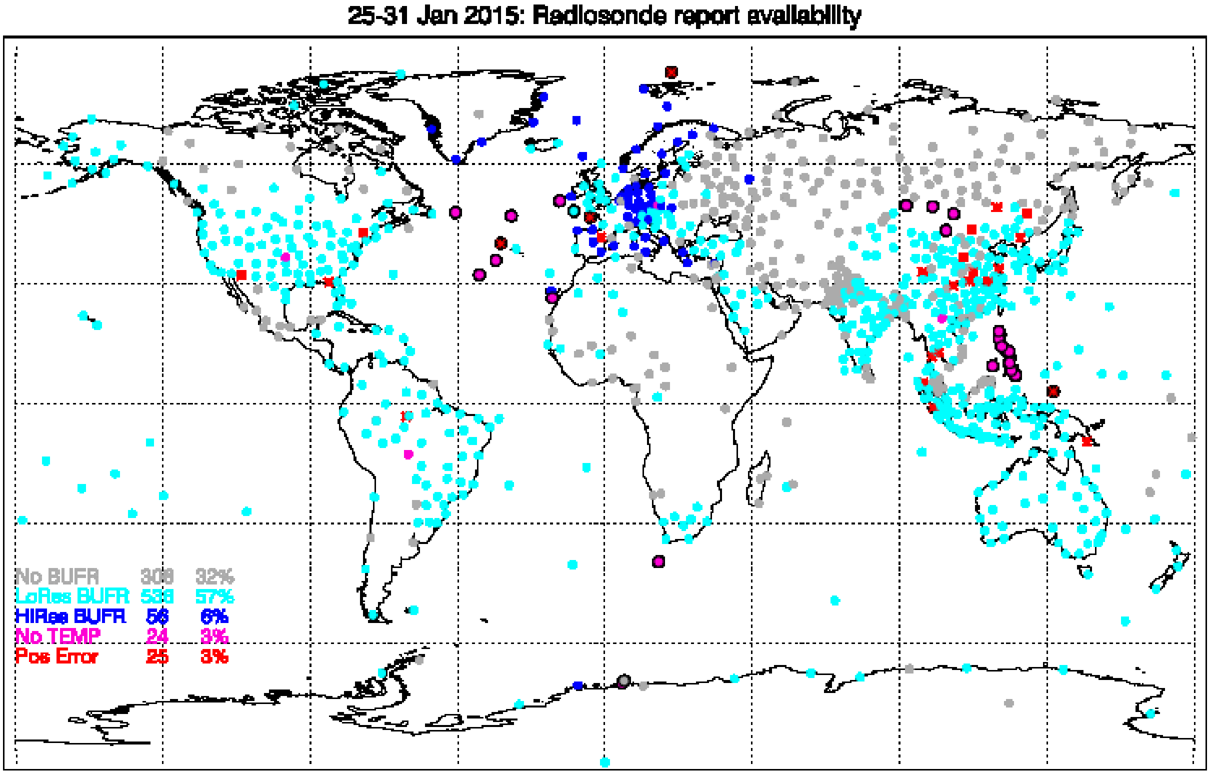

Figure 2. Availability of radiosonde reports at ECMWF for 25-31 January 2015. Grey – no BUFR, blue - both TAC and BUFR (light/dark blue – low/high resolution available in BUFR), purple – no TAC. Red symbols indicate minor position errors (differences from WMO publication 9A) – large position errors have been corrected. A black ring around the station indicates the absence of a numeric WMO block/station number: either for ships or stations from Mongolia and the Philippines which have a metadata error. (Reports from Mexican stations were not decoded at the time due to the use of an obsolete BUFR edition.)

Figure 2 shows radiosonde report availability for the same period (21-25 January 2015). As for land surface data almost 70% of stations are reporting some BUFR. The spatial coverage is similar, but only about half of Canadian radiosonde stations are reporting BUFR and only two Russian stations are. In some cases sections of ascents are missing (e.g. upper level temperatures – above 100 hPa – are missing from Portuguese and South African stations). Some problems with station position occur (the same issues as for SYNOP, but most of the largest position errors have now been corrected for radiosonde reports). China has a number of position errors (differences from WMO publication 9A) and also some stations not reporting BUFR.

Currently ECMWF does not store the time of receipt in the observation feedback database (but this is planned), so we can say very little about timeliness. It seems that Mexican BUFR radiosonde data is relatively late as it does not make the NAVGEM cut-off (Pat Pauley, Naval Research Laboratory, California, USA).

Radiosonde reports from Mongolia and the Philippines have missing WMO identifiers (the ECMWF system uses the reported station names as a character identifier and labels them as ship reports, the Met Office sets blank identifiers. Dropsondes currently have blank identifiers in BUFR. Some ships making radiosonde ascents changed identifier slightly (e.g. from ASDE1 to ASDE01).

Extra radiosonde metadata was discussed along with BUFR coding issues (section 2). Radiosonde type uses code figures 0-99 in TEMP and 0-255 in BUFR and some types have 100 added in the BUFR version of the table. A few stations report Vaisala RS41 as type 41 in both TAC and BUFR – it should be 141 in BUFR.

In November 2014 TAC TEMP reports from some ships (ASDE* and ASEU* ASAPs, also DBLK) ceased to be circulated. In January 2015 notice was given that TAC TEMP reports from the Netherlands will cease.

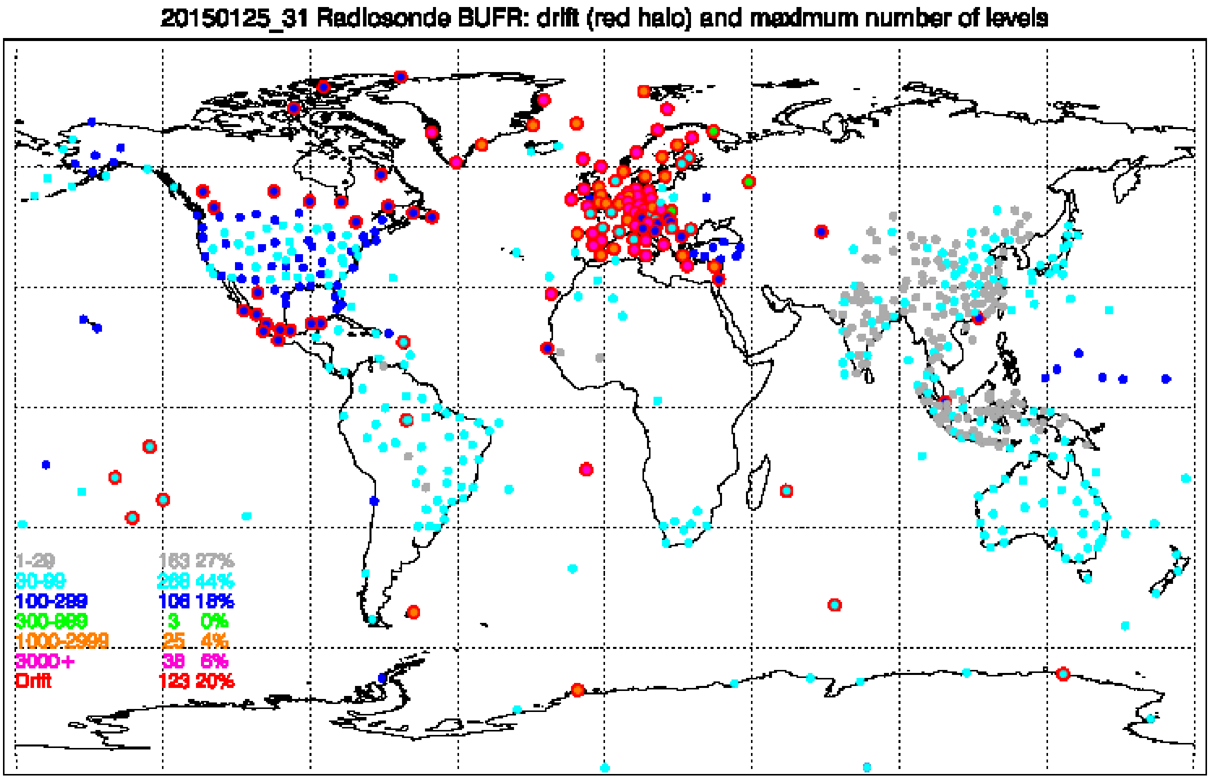

Figure 3. Details of BUFR radiosonde reports decoded at Met Office for 25-31 January 2015. A red ring indicates valid drift positions. The colour of the main dot indicates the maximum number of levels reported: Grey – 1-29, light blue – 30-99, dark blue – 100-299, green – 300-999, orange – 1000-3999, purple – 3000+.

5.2 Station generated (native) BUFR radiosonde reports

There is no single clear way to distinguish native BUFR from reformatted TEMP but the easiest test in practice is to check for non-zero displacement (drift) positions, see Figure 3. Of those stations reporting valid displacement positions on 25 January 2015 approximately 70% use Vaisala RS92 radiosondes, 12% Modem M10, 10% Graw, 4% Intermet, 2% Russian, 2% Vaisala RS41. (Note that many Indian radiosondes report zero displacements at all levels and the data appears to be reformatted TEMP.) The Modem data is relatively low resolution (~70 levels) and some ascents show extra noise in some wind levels compared to TAC (for part of 2014 there were some occurrences of negative dew point depression, but these have been resolved). The wind noise has been discussed with Météo-France and we understand that at some point this year Modem will enable higher resolution reporting – ECMWF intend to wait for this before re-evaluating the quality of the Modem BUFR reports and (hopefully) making them operational. The other native BUFR reports generally look to be of good quality (with two exceptions: 1. Mexican reports use an obsolete edition of BUFR and did not reach the ECMWF NWP system for this reason, 2. Reports have been seen from three stations in Greece and Cyprus with a low resolution version of the ascent appended to the high resolution version).

Many stations using Vaisala radiosondes (including the ships) send low resolution reports (essentially standard plus significant levels) from the surface to 100 hPa but send high resolution full reports (e.g., levels every 10 seconds from ASAP ships, limited by communication costs – giving typically 500 levels, every 2 seconds from many European land stations – giving 3000 levels or more, a few reports have about 8500 levels). Canadian (RS92) and Mexican (Graw) BUFR reports are typically between 100 and 170 levels. (The main Canadian reports are good quality BUFR, but the winds are also sent as separate BUFR reports causing some duplication/confusion.) We understand that in reports from Vaisala RS92 dew point depressions (the difference from air temperature) above 49° aren't coded in TEMP (depressions reset to 49°?) but depressions up to 65° are coded in BUFR. The (high resolution) Modem reports from 04360 (Greenland) did have a wind scaling problem – now fixed, but the reports have various minor problems: some levels labelled standard that aren't, some levels out of order and some gross errors in longitude displacement. Two Finnish stations using Vaisala radiosondes are producing low resolution reports.

Radiosonde 'time of report' usually differs between TAC and BUFR. Typically the balloon is launched about 45 minutes before the main synoptic hour (say 1115 UTC for 1200 UTC nominal ascent, but sometimes 1030 UTC or earlier). The BUFR report gives the launch time in hours and minutes (and sometimes seconds) whereas the TEMP report gives it in hours (often 1100 UTC for the case above, but sometimes as 1200 UTC) complicating report matching. (Generally there is only one ascent per synoptic hour, but there may be two if the first ascent terminates prematurely; also there can occasionally be frequent ascents for field experiments).

There can be small (~0.1°) temperature differences between TEMP and BUFR reports (see Ingleby and Edwards, 2015; TEMP reporting is in 0.2° intervals in Celsius and there are issues of rounding and conversion to Kelvin). These differences are at the noise level for NWP but could be a problem for climate trend statistics. Some of the reformatted TEMP reports (see Section 5.3 below) are TEMP-like in their rounding, whereas others are BUFR-like in their rounding – causing extra problems for climate users unless the TAC and native BUFR reports overlap.

5.3 Reformatted TEMP reports

Generally reformatted TEMP reports are rather messier than native BUFR reports. There are typically 4 reports (parts A, B, C and D) but there can be more; and there can be 8 or 10 reports where the winds have been separated (including cases with inconsistent wind vertical coordinates as discussed for TAC). Levels appearing out of order are not uncommon. In some cases different parts of the same ascent are given different report times. Some reports/parts 'go missing' at various stages (it can be difficult/time-consuming to determine where/why). At ECMWF upper levels (above 100 hPa) of South Korean ascents are missing from BUFR – at the Met Office the whole ascents are missing from BUFR. For a small proportion of US ascents ECMWF only decodes one report/part whereas the Met Office decodes four.

Several users have noted problems with Indian stations reporting dew point depression instead of dew point temperature (not all the time but quite frequently). Alexander Kats (Central Aerological Observatory, NTCR, Russia) noted (email, Jan 2015) various other problems with Indian BUFR reports. Both India and China typically only report about 30 levels in TEMP or BUFR.

Errors that are still awaiting correction (noted by Pat Pauley):

- Dewpoint depressions between 5°C and 6°C appearing in BUFR as values between 50°C and 60°C (Thailand)
- Problems with the reported tropopause level (India and Thailand)
- Significant level winds above 100 hPa for some stations in India with pressures divided by 10.
- Surface winds repeated as upper level significant winds (Oman)
- 925 hPa level converted to 920 hPa (Vietnam)
- Surface data repeated as 1000 hPa data (Australia, where 1000 hPa level is sub-surface)
- Significant levels below 1000 hPa assigned pressures without the leading 1 digit (Japan)
- Incorrect launch time, wrong leading digit for 7 hPa and 5 hPa geopotential height and for 250 hPa and 10 hPa when heights are low, missing tropopause interpreted as having a pressure of 999 hPa (U.S.—legacy format)
- Metadata errors (U.S. and other countries)

New Zealand generates four pressure/temperature reports and four wind-only reports both in TAC and BUFR, the wind-only reports use different vertical coordinates as discussed above (valid in TAC but merging the parts back together is messy as a result). A Kats noted two minor coding errors: a) 0-08-042 in IUK/IUS ii=01 bulletins converted from New Zealand TEMPs parts A/C should not have bit 6 set to 1 - significant humidity level, it's wrong, b) 0-08-042 in IUW/IUJ ii=01 bulletins converted from New Zealand PILOTs parts A/C should have bit 17 set to 1 - pressure level originally indicated by height as the vertical coordinate.

In early 2015 some significant level wind speed reports from Vietnam were too weak by a factor of about 2 (reports in km/hour? Or converted when they didn't need to be?).

In order to limit values to three or four digits in TEMP code sometimes the leading digit is omitted and has to be inferred. The TAC decoders have been used over many years and include special cases (or tricks) to correctly infer the leading digit – not all of these special cases are correctly treated in TAC to BUFR conversion programs. "A few problems were found in TEMP BUFR reports sent from Tokyo. When the geopotential height of 700 hPa goes below 2500 m, it is reported 1000 m higher than actual. When there are significant levels below or equal to 1000 hPa level (i.e. pressure >=

1000), the pressure is reported 1000 hPa less than actual." There are also problems with negative heights. (Eizi Toyoda, Japan Meteorological Agency) *More checks on geopotential heights?*

For 26 January 2015 some basic sanity checks were made for:

- temperature less than 170 K: India (5 stations had at least one level)
- temperature greater than 325 K: Vietnam (1 station)
- missing vertical coordinate: Japan (2 stations), Thailand (1), Vietnam (2), Cook Islands (1), Philippines (7), Mongolia (4)

It is quite likely that a different decoder would show either more or fewer errors in the case of format errors. All of these countries currently transmit reformatted TEMP.

In January 2015 the ECMWF routine observation monitoring was extended to BUFR radiosonde data and identified stations 42182, 43003, ASDE00, ASEU00 as having suspect Geopotential Height and 48407 and 78073 as having suspect Wind statistics (among these, 42182, 43003 and 48407 were also on the corresponding TAC suspect lists).

6. Use of BUFR data in NWP

On 16 January 2015 a quick survey was sent out to global NWP centres, the results are summarised in table 1. For BUOY the Météo-France response "waiting for data producers to use the new WMO templates" also applies to ECMWF and probably other centres. So far only Météo-France is assimilating BUFR ship data (ECMWF is processing the data, but has only recently started looking at the data). The proportion of BUFR land surface data used varies widely. Many centres are not using BUFR radiosonde data yet, ASAP (ship) data is being given priority by those that are. The percentages in Table 1 are approximate.

Centre	TEMP/PILOT	SYNOP	SHIP	BUOY	Notes
Australia, BoM	-	-	-	-	
Canada, CMC	0.5%	2.5%	0%	0%	TAC used if available
China, CMA					Status unknown
Europe, ECMWF	2%	2.5%	0%	-	
France, MF	-	70%	85%	-	SYNOP/SHIP BUFR used if available
Germany, DWD	0.5%	25%	0%	-	
Japan, JMA	0.5%	2.5%	-	-	TAC used if available
Korea, KMA					Status unknown
Russia, RHMS	-	-	-	-	
UK, Met Office	-	1%	-	-	(UK BUFR ascents used in UKV)
US, FNMOC	-	-	-	-	
US, NCEP	-	-	-	-	

Table 1. Approximate proportion of actively assimilated reports provided by BUFR, Jan 2015. The questions asked were: "Are BUFR reports of this type processed in your operational Global NWP system? If Yes, then what proportion of active assimilated reports currently come from BUFR (e.g., 10% BUFR versus 90% TAC). If No, when do you expect to process BUFR reports of this type. Any notes/comments?" Some centres are reformatting BUFR surface reports back to TAC SYNOP format

before ingesting it into their systems (this can make it more difficult to distinguish the GTS format used). A few centres are still receiving UK TAC SYNOPs.

Given the November 2014 "deadline" why aren't all NWP centres using the BUFR data? There are various possible reasons: uncertainty/unclear information about the migration (previous "deadlines" passed without effect) and the shifting nature of the BUFR data available, pressure of other work combined with little perceived benefit of BUFR, operational upgrade cycles (generally one or two upgrades a year, which can slip), the need for different groups within each centre (database/decoding and data assimilation) to be involved. (One correspondent suggested that oceanographic users are even less prepared for the migration.)

If they don't use BUFR data then NWP centres risk increasing data gaps giving slightly worse forecasts. However there is also a more acute risk if the assimilation of poor quality data is attempted – in early 2014 the Japanese were assimilating some BUFR radiosonde data when the US introduced a reformatting bug causing significant bias of the geopotential height - and a distrust of reformatted TEMP data at JMA (E Toyoda). In January 2015 for a few days all the US BUFR radiosonde reports had the wrong sign of longitude – ECMWF has put measures in place to guard against position errors, but other more subtle errors could still cause serious problems.

In the ECMWF NWP both TAC and BUFR reports are processed (not BUFR buoy data yet) so that they can be compared and monitored. Initially all the BUFR reports were 'blacklisted' but when a particular subset of BUFR appears stable and of good quality it has been activated operationally. The subsets are usually stations from a particular country – sometimes when that country has announced the cessation of its TAC reports. For SHIP data the distinction between countries is less clear cut (and GTS bulletin headers are not stored in the NWP observation data base), so it is likely that BUFR ship data will be activated all at once (and let the thinning routines choose between TAC and BUFR reports where both occur).

The table above represents usage in the main atmospheric analysis: for global systems the main surface variable of interest is pressure although use of other variables is increasing (Ingleby, 2014). At ECMWF there is also a) a snow depth analysis, b) a surface and soil moisture analysis (which uses screen temperature and humidity), c) the ERA-Interim real-time extension and d) an atmospheric composition analysis (which also includes all the conventional variables such as air temperature). These are not yet using BUFR data although work is underway to incorporate them. There are also changes needed to observation monitoring and verification, some of which have started.

The displacement positions in native BUFR radiosonde data are potentially useful in taking account of balloon drift (Laroche and Sarrazin, 2013; Ingleby and Edwards, 2015), but care will be needed to exclude erroneous positions. High resolution reporting will also provide improvements to NWP (more for some systems than others). ECMWF thins high resolution reports to about 350 levels per ascent, the Met Office averages the ascent over model layers. The large number of levels per report (with no fixed upper bound) may necessitate various code changes (it was recently found that ECMWF wasn't decoding reports with more than 6000 levels – this is now being fixed).

6.1 Non-regulation BUFR radiosonde data in parts

This has been discussed from coding and quality perspectives above (sections 2 and 5). For TEMP (and PILOT) the Met Office, JMA, NCEP and ECMWF all merge the different parts to form a single ascent. This is not straightforward: different parts can have slightly different times, the levels have to be sorted into order and it is difficult to code efficiently and robustly in a parallel computing environment (different parts are handled by different processors). The coding is specific to alphanumeric TEMP reports – the centres would like to retire this task, not to recode it for non-regulation BUFR reports. The reasons for merging the TEMP parts are:

- For vertical consistency checks
- For vertical averaging (Met Office, see Ingleby and Edwards, 2015)
- So that the radiosonde type is available for the whole ascent

For NWP the radiosonde type is used in several ways at ECMWF and/or the Met Office: a) to determine whether to use upper tropospheric humidity measurements and stratospheric temperature measurements, b) radiation bias correction, c) observation error estimates. In TAC the radiosonde type is only available in Part A and this is generally the case for reformatted TEMP too. To use reformatted TEMP in the Met Office NWP system would require code merging the different parts. ECMWF might avoid merging BUFR reports in parts but would have to copy the radiosonde type to all related parts (unless the data producers do this) and needs to distinguish native BUFR from reformatted TEMP because different vertical thinning algorithms are used. Similar issues are likely for other NWP centres. As discussed above there are small temperature offsets (~0.1°) between TEMP and native BUFR, these are too small to affect NWP much but may be an issue for reanalysis and climate studies; reformatted TEMP sometimes takes rounding into account and sometimes doesn't.

There was a comment from JMA (Japan) about the inconsistency in styles of reporting and the errors that occur in reformatted TEMP data. The data assimilation section at NCEP (US) was concerned about the difficulty of creating a full profile from converted TEMP parts and preferred to wait until the reports are available as native high resolution BUFR.

There is also a lesson from history (pointed out by Pat Pauley). In Eugenia Kalnay's 2002 book "Atmospheric Modeling, Data Assimilation and Predictability", she describes what happened at NCEP in the 70's when they switched to Office Note 29 (a predecessor to BUFR). "This change in formatting required a complete overhaul of the NMC decoding system and errors must have been introduced during this complex reprogramming process. The NMC operational forecast skill actually went down and it took a few years before it recovered to the pre-1974 error levels."

In general operational NWP centres only make code changes with major upgrades (once or twice a year), changes will only be made at other times *in extremis*. ECMWF, the Met Office and probably other centres have "blacklist/whitelist" control files (changed more frequently) which can be used for example to assimilate BUFR data from an extra country – provided that this works with existing code.

In summary native BUFR (whole ascent in a single report) provides a much cleaner interface than TEMP for NWP. It would be a backward step if users have to assimilate non-regulation BUFR data in parts as well as TEMP and native BUFR data.

7. Migration procedure

Notification and checking of changes

Notification of the cessation of TAC reports from a given subset of stations (usually a particular country) can be made via various ways. A) WMO operational newsletter (plain text), B) METNOs (see below), C) WIS (being spun up, no notifications seen this way yet) and D) emails. To date most notifications have been via METNOs – a slightly cryptic format which until now was of no interest to Data Assimilation practitioners. Sometimes METNOs are sent with no advance notice or an unclear implementation date. A major use of METNOs is to change GTS routing (GENOTs are also involved, in 2014 some Canadian TAC radiosonde data was missing from various NWP centres because of an unclear GENOT). On 18 December 2014 New Zealand stopped sending TAC data having given a few days' notice via METNO (after requests they restarted sending TAC data two days later). The timing was unfortunate as some NWP centres have a moratorium on changes in late December because of holidays. Also at issue was the fact that New Zealand was (and still is) sending reformatted TEMP data still as parts (as discussed above).

Even when good care is taken errors can occur when using a new format (as with BUFR). Finding errors in a new data stream is much easier if the same observations are available in both old and new format. For both reasons we **need overlap** (typically for several months) between TAC and BUFR: to assess the availability and quality of the new data and either a) report problems to the data producer or b) make a change to assimilate the new data. Also users **need notice** that the data producers regard a particular BUFR subset as usable/finalised and intend to cease transmitting the TAC version (in, say, two months' time). If, as some have proposed, reformatted TEMP data still as separate parts is allowed as an interim measure then this increases the total work needed in assessing new data streams. Also if there is a change from reformatted TEMP (in BUFR) to native BUFR and they are transmitted in parallel then users need to distinguish these (and blacklist native BUFR until it has been checked). Effectively there would be three data formats and variants of the processing: TAC, reformatted TEMP and native BUFR. It is much simpler from an NWP viewpoint to have a single change from TAC to native BUFR (continuing to use a decreasing proportion of TAC reports for several years is much less of a problem for NWP users than extra transitions and forced use of reformatted TEMP data still in parts). *Reference the draft change procedure document from Steve Foreman?*

Radiosonde manufacturer's processing software changes infrequently (usually once every few years, perhaps more frequently during the BUFR migration) but reformatting programs can change more often (with serious consequences if a mistake is made, discussed above). Ideally NWP centres would want to wait until the reformatting program has been finalised before (re)checking a particular subset of BUFR reports and making them operational.

8. Summary and prospects

The migration has involved a lot of work from a lot of people over the years but unfortunately is still far from complete. The users (notably NWP centres) have only become involved relatively recently but some are now putting significant effort into decoding the BUFR data, reporting any quality problems and using subsets of the BUFR data. NWP centres are naturally cautious when it comes to changing operational systems and need plenty of notice of the withdrawal of subsets of TAC data – this notice has been lacking with some changes. Some NWP centres now have small gaps in data coverage because they are not yet ready to use the BUFR reports. Arguably data users should have been more involved in planning the migration.

There are errors in **BUFR land surface** reports from a few countries, and still some position errors to resolve, but most reports are usable and the main issue is that BUFR reports are not available from about 30% of stations. For some countries the BUFR land surface reports are less frequent than the TAC SYNOP reports (e.g. six hourly, rather than three hourly). There are (near-)duplicate BUFR land surface reports from some countries, sometimes in two different templates. **BUFR ship** reports have received less attention to date but (with one minor exception) appear to have been correctly reformatted from TAC (around 10% of TAC reports do not appear in BUFR). The main concern is that the current BUFR marine templates are a stop gap measure (and ship, moored buoy and drifting buoy reports are rather mixed together, as with TAC) and there is another transition to newer templates in the near future. ECMWF would rather wait for the new templates before assimilating BUFR ship and buoy data. There is also a question about the readiness of oceanographic users for the cessation of TAC data.[1]

For surface data reformatting TAC to BUFR is common and generally successful (although it doesn't take advantage of the extra precision and metadata of BUFR). For **radiosonde** data reformatting TEMP/PILOT reports is more problematic, with reformatted reports usually still as separate parts (in violation of BUFR regulations) and also more subject to reformatting errors. Proposals to allow BUFR radiosonde reports that do not comply with the BUFR coding regulations would add to the work needed for NWP and, if not very carefully handled, adversely affect the quality or quantity of the data assimilated.

Radiosonde manufacturers provide options to produce BUFR reports in addition to (or instead of) TAC reports. The BUFR reports are often high vertical resolution are generally good quality. The main problem is that currently such native BUFR reports are only available from about 15% of stations (mainly from Europe). However the proportion of native BUFR reports is expected to increase significantly over the next year: Australia and the US are currently working on high resolution BUFR radiosonde reports (whilst sending reformatted TEMP for now), Russia and Canada intend to produce native BUFR from the balance of their stations (not currently available in BUFR). The plans of most other countries are not known to the authors, and we suspect that some do not realise that their reformatted reports do not meet the BUFR regulations. Some countries have no spare resources to make changes to their radiosonde programme and will presumably need help via voluntary cooperation programmes.

It seems likely that it will take years, rather than months, until the migration is complete and that radiosonde migration will take longer than the surface data types.

Acknowledgements

Only a subset of the many people from many different NWP centres that have contributed in one way or another to this report have been mentioned in the text. We acknowledge all of them.

[1] In parallel there is also a move towards BUFR for aircraft (AMDAR) reports. There are fewer data producers involved making the change somewhat simpler.

References

Ingleby B., 2014: Global assimilation of air temperature, humidity, wind and pressure from surface stations, Q.J.R. Meteorol. Soc. doi: 10.1002/qj.2372

Ingleby B. and D. Edwards, 2015: Changes to radiosonde reports and their processing for numerical weather prediction, Atmosph. Sci. Lett., 16: 44-49. doi: 10.1002/asl2.518

Laroche S, Sarrazin R. 2013. Impact of radiosonde balloon drift on numerical weather prediction and verification. Weather and Forecasting 28: 772–782, DOI: 10.1175/WAF-D-12-00114.1.

Appendix 1. Further information

A summary of current BUFR availability and quality (with contributions from various countries) can be found at https://software.ecmwf.int/wiki/display/TCBUF/ . Information on surface marine reports is at https://software.ecmwf.int/wiki/display/TCBUF/E-SURFMAR .

BUFR templates and regulations (e.g. B/C20 and B/C25 for PILOT and TEMP):
http://www.wmo.int/pages/prog/www/WMOCodes/TemplateExamples.html

"radiosonde/sounding system used" is defined in common code table C-2 of the WMO Manual on Codes Volume I.2 available from http://www.wmo.int/pages/prog/www/WMOCodes.html

Details of radiosonde height coding and errors at http://toyoda-eizi.blogspot.co.uk/2013/12/ambiguity-in-pressure-level-heights-of.html

Appendix 2. Miscellaneous problems

Radiosonde stations (from Mongolia and Philippines) with **missing WMO block and identifier** in BUFR: 'Baguio' 'Dalanzad' 'DavaoAi' 'Laoag' 'Legaspi' 'LumbiaA' 'Mactan' 'Muren' 'PuertoP' 'Tanay' 'Ulaan-Ba' 'Ulaan-Go'

BUFR radiosonde reports with position errors (differences from WMO Publication 9A; # denotes larger differences). *Data from Met Office, in some cases the difference is from SYNOP position.*

01415 02591 04270 04692# 08160# 11120 15420# 16716# 17064# 42103# 42273# 42369# 42397# 42623# 42634# 42867# 42874# 43049# 43185 43285# 47418# 48354 48431 48480# 48500# 48565 48698# 50527# 50774# 50953# 53463# 53513 53845# 54218# 54374# 54511# 56029# 56187# 57494# 57516# 57816# 58238# 58362 58424# 59280 59316# 60018# 60571# 61641 61687 61901# 70026 70414# 71203 71603# 72214# 72215# 72274# 72293 72388# 72403 72426 72493 72662# 72672 72681 74005# 74646# 76654 76679# 78762 81729# 82026# 82107# 82193# 82244# 82281# 82397# 82411# 82532# 82705# 83208# 83362# 83378# 83525# 83554# 83566# 83612# 83649# 83746# 83779# 83899# 83928# 83937# 83971 85442 85934# 88889# 89022# 92035# 96009# 96075# 96171# 96295# 96645# 97240# 97300# 97340# 97372# 97430# 97460# 97502# 97530# 97560# 97600# 97686# 97690# 97724# 97748# 97760# 97810# 97900# 97980#

Add the following?

Zstn errors/differences. Missing/incorrect radiosonde type? BUFR surface position errors. (long list)