Assimilating Real-Time High-Resolution BUFR Radiosonde Data in NAVGEM: Current Results and Future Prospects

Patricia Pauley and Daniel Tyndall
Marine Meteorology Division, Naval Research Laboratory, Monterey CA
Justin Reeves and Randal Pauley
Fleet Numerical Meteorology and Oceanography Center, Monterey CA

Why use BUFR for radiosonde data?

TAC: Traditional Alphanumeric Codes

- Developed for the teletype era but still in use
- "TEMP" code used for radiosonde data
  - Each sounding has four parts (TTAA, TTBB, TTCC, TTDD)
  - Parts A and C—mandatory levels below/above 100 hPa
    - Pressure, height, temperature, dewpoint depression, winds
  - Parts B and D—significant levels, below/above 100 hPa
    - Pressure, temperature, dewpoint depression
  - Precision limited and variable
- "PILOT" code used for winds (pilot balloons, etc.)
  - Parts A, B, C, D like TEMP (PPAA, PPBB, PPCC, PPDD)
  - Parts B and D have geopotential height, winds
  - Used in some countries (e.g., U.S.) for significant level radiosonde winds

```
TTAA 56121 72662 99892 12262 27505 00047 ///// ///// 92717
///// ///// 85430 11263 32026 70018 00058 32553 50563 13557
32078 40729 25560 33077 30931 41356 34104 25052 51159 35120
20195 58557 35127 15374 59362 33588 10628 59368 34058 88181
```

Shortcomings of TAC:

- Knowledge of arcane "rules" is required to read reports.
- Changes such as increasing precision and adding new quantities are impossible to make.
- Metadata (latitude, longitude, elevation, etc.) are typically not included, requiring an external station list.
- Four to six files are required to reconstitute a sounding.
- If significant level winds are provided in PILOT rather than TEMP messages, the geopotential heights need to be converted to pressure, which is a significant source of error.

BUFR: Binary Universal Form for the Representation of meteorological data

- First approved for use by the WMO in 1988
- Binary—files are compact
- Universal—one decoder used for many data types
- Self-describing—through use of "data descriptors"
- Expandable—easy to add variables, precision
- Flexible—easy to vary content
- Sustainable—decoders backward compatible
- WMO has approved "templates" for specific data types (e.g., radiosonde data) with lists of defined descriptors to make it easier to encode data
  - Precision is improved over TAC.
  - Includes descriptors for station metadata so an external station list is no longer necessary.
  - Thousands of vertical levels can be included.
  - Radiosonde drift time, lat, lon offsets are included.
  - Two messages per sounding are intended—one when the balloon reaches 100 hPa, the second with the full sounding from the surface to balloon burst.

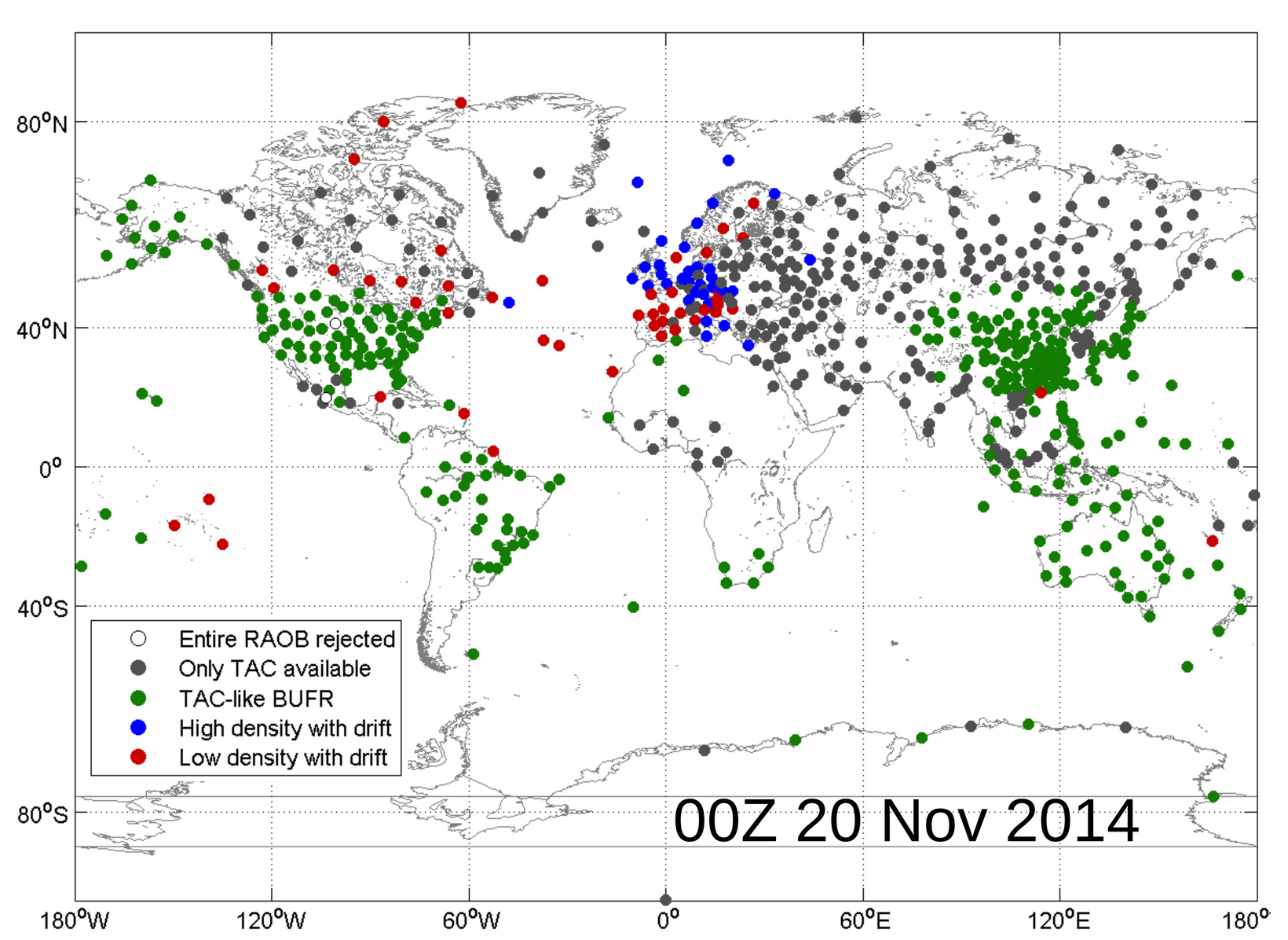
```
043930600372600 1250999999901100004101020103312264001710110
40340025AA A01503000003034002 4AA A03039-071160340033AA A0558
0-211999320048AA A07180-351130320060AA A09 40-461999310061AA
A10340-311140300061AA A11790-521999310047AA A13640-54199931
```

Shortcomings of BUFR:

- Not human readable
- Requires considerable expertise to implement

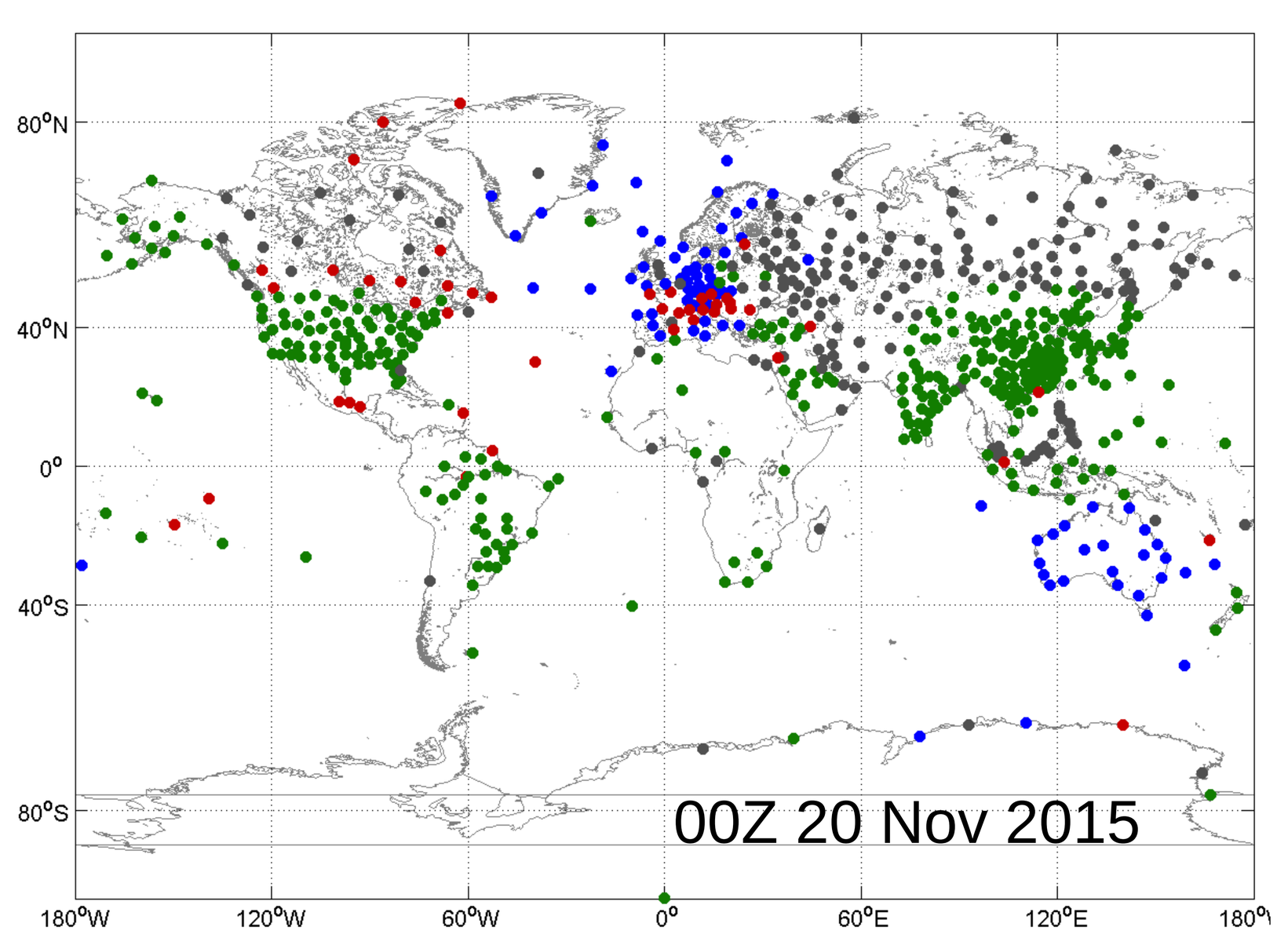
The TAC to BUFR Migration

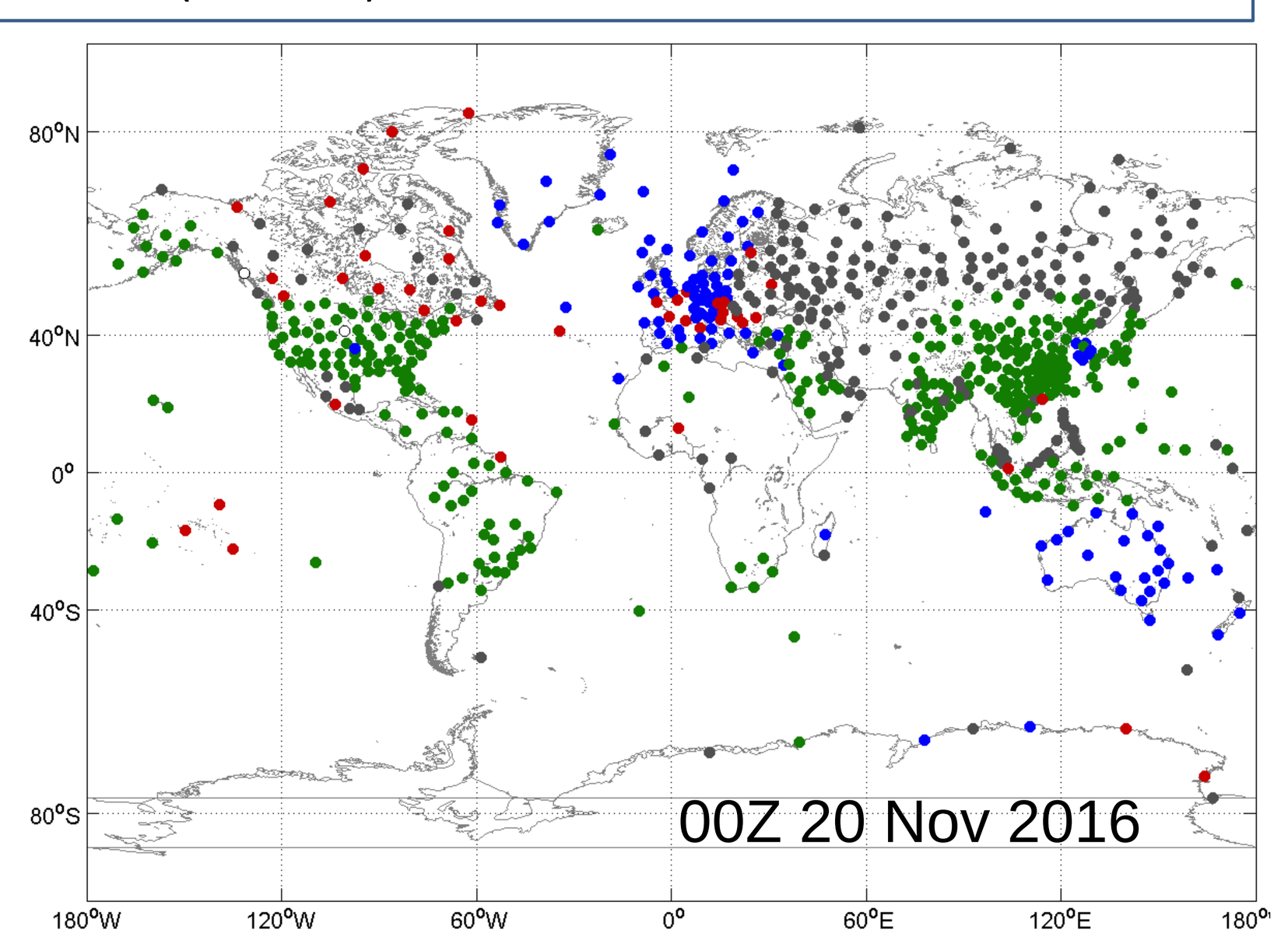
- In 2003, the WMO members approved a migration from TAC to BUFR for data distribution on the Global Telecommunications System (GTS), the primary communications pathway countries use to share data.
- Data were to be encoded and distributed in both TAC and BUFR for a limited time to enable users to transition between the two formats.
- The official timetable called for TAC distribution to cease in November 2014, however not all countries have started producing BUFR and only a small fraction of TAC has actually stopped.



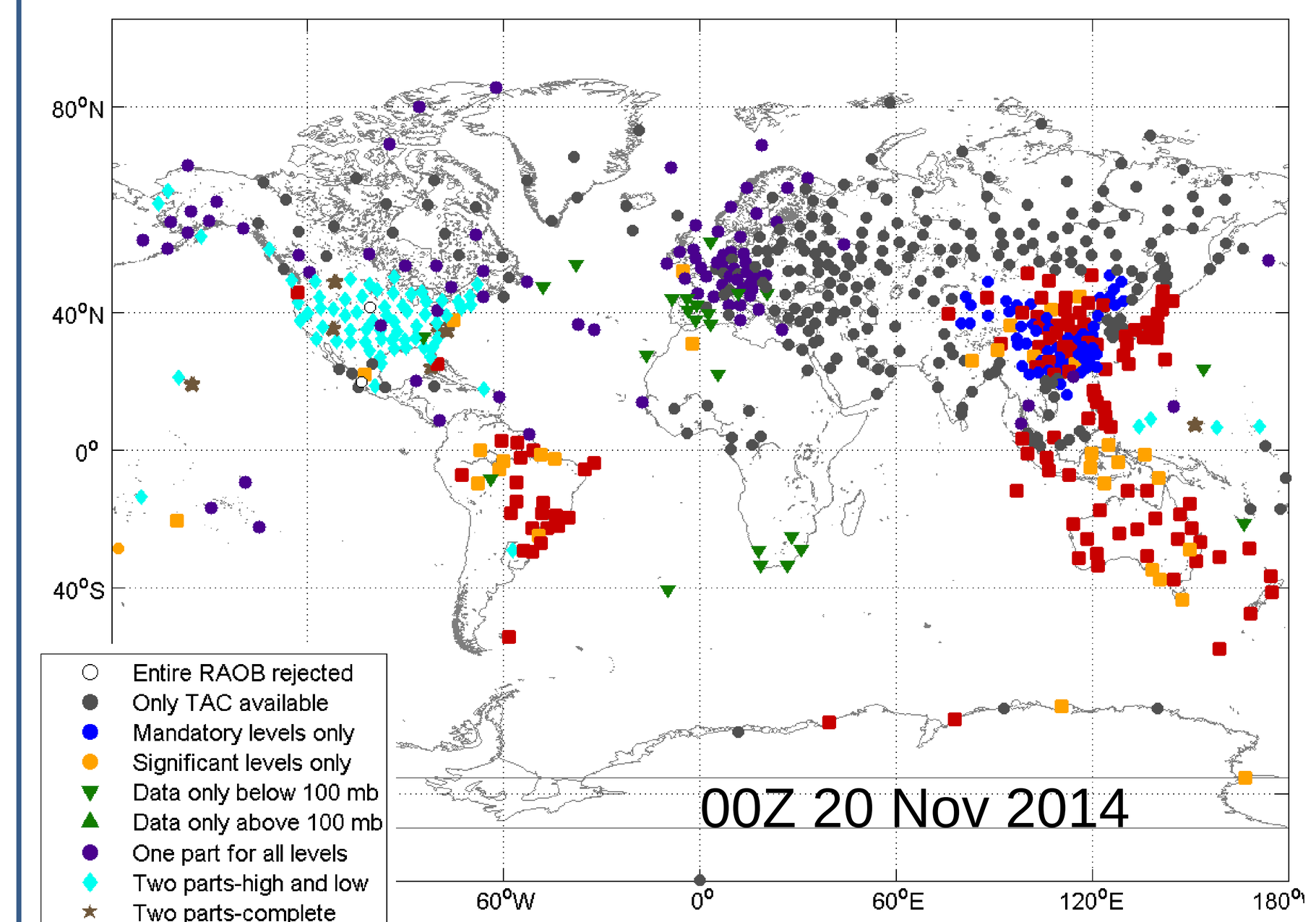
"Native" vs. reformatted BUFR

- "Native" BUFR data use data from the radiosonde ground-station and include balloon drift information.
  - High-density native BUFR data have hundreds to thousands of levels (blue).
  - Low-density native BUFR data have a couple hundred levels (red).
- Reformatted BUFR use data decoded from TAC bulletins (green) and so do not have drift data.
  - Reformatted BUFR have the same number of levels as TAC (~60-100).



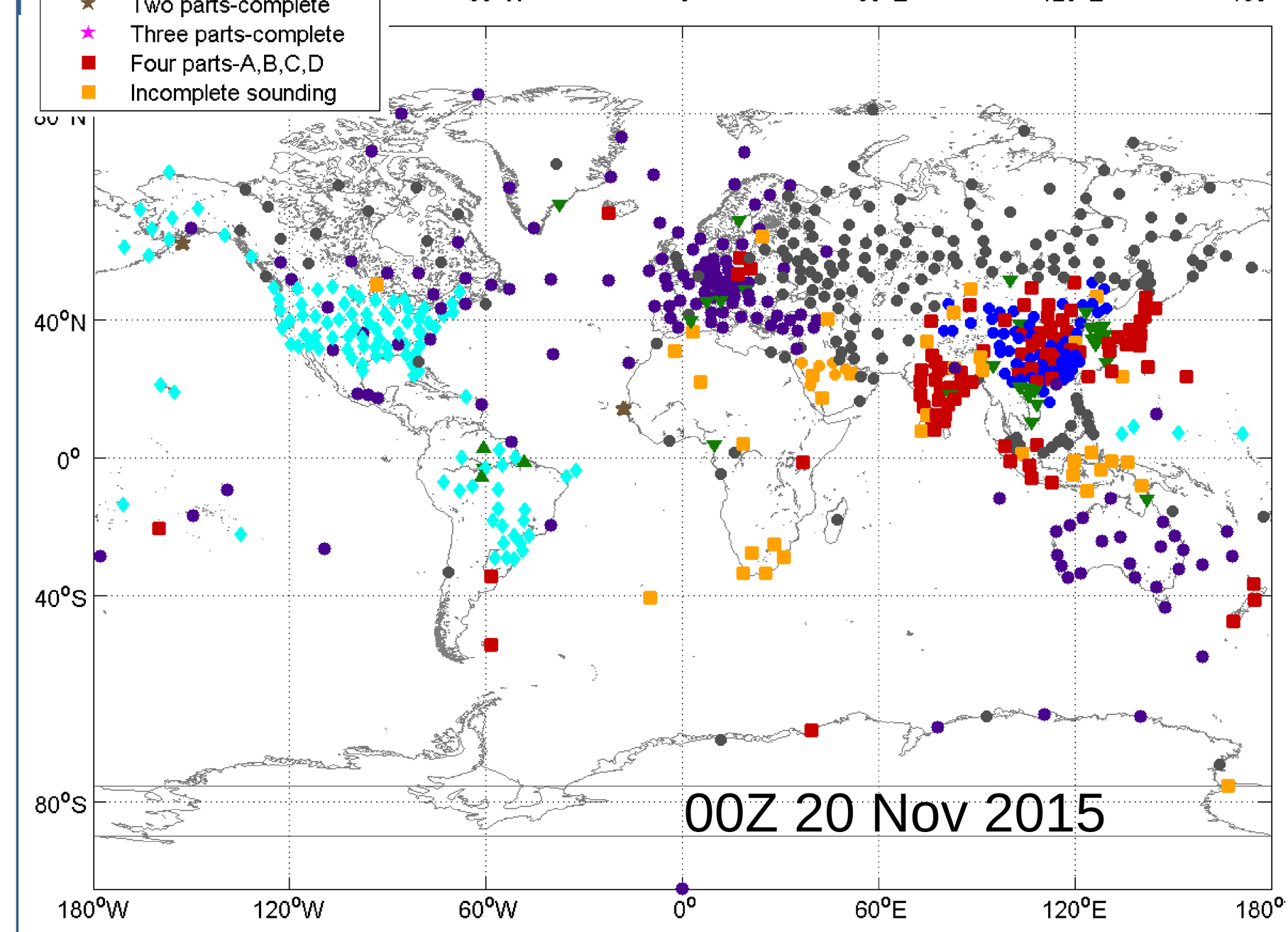


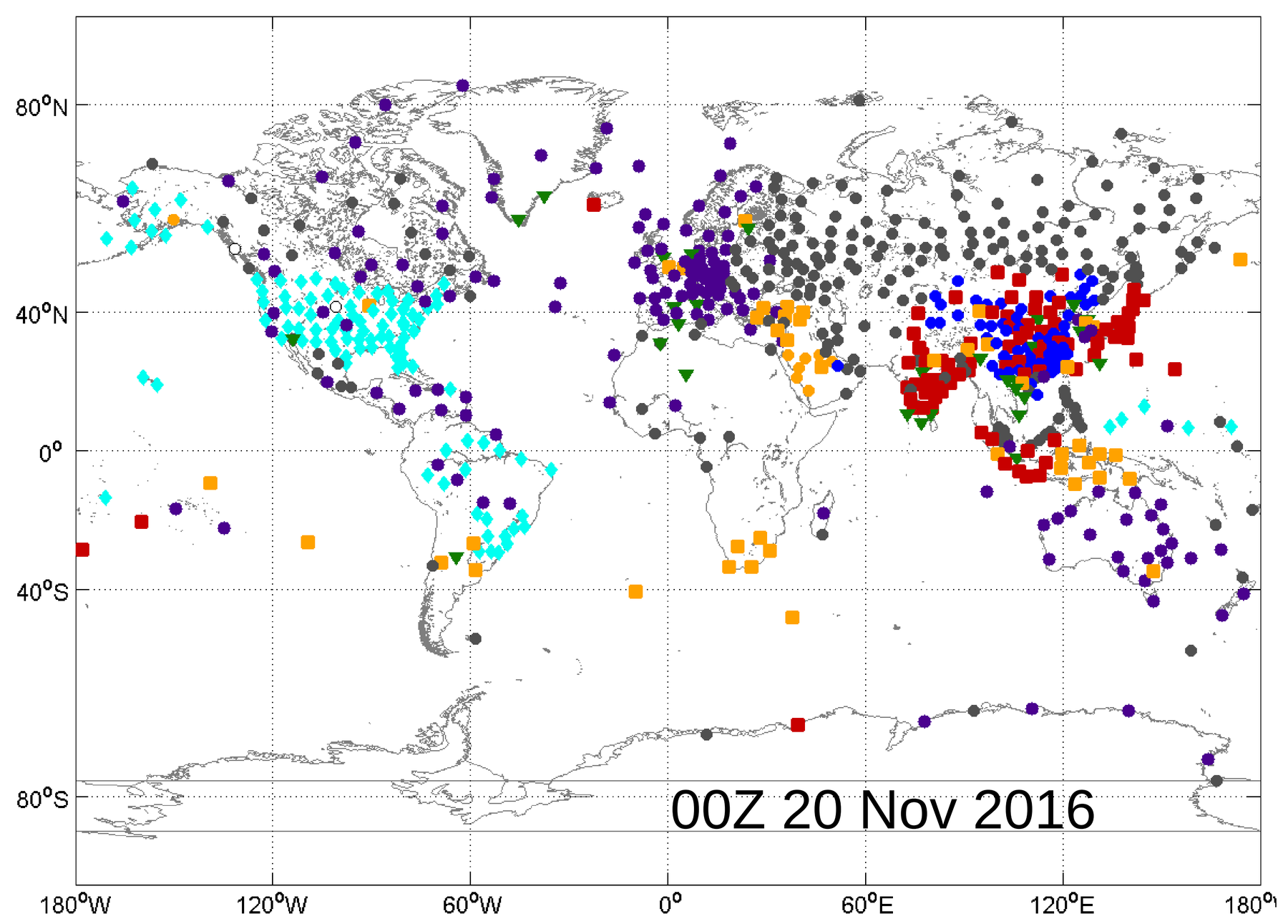
Characteristics of BUFR radiosonde data



BUFR structure issues

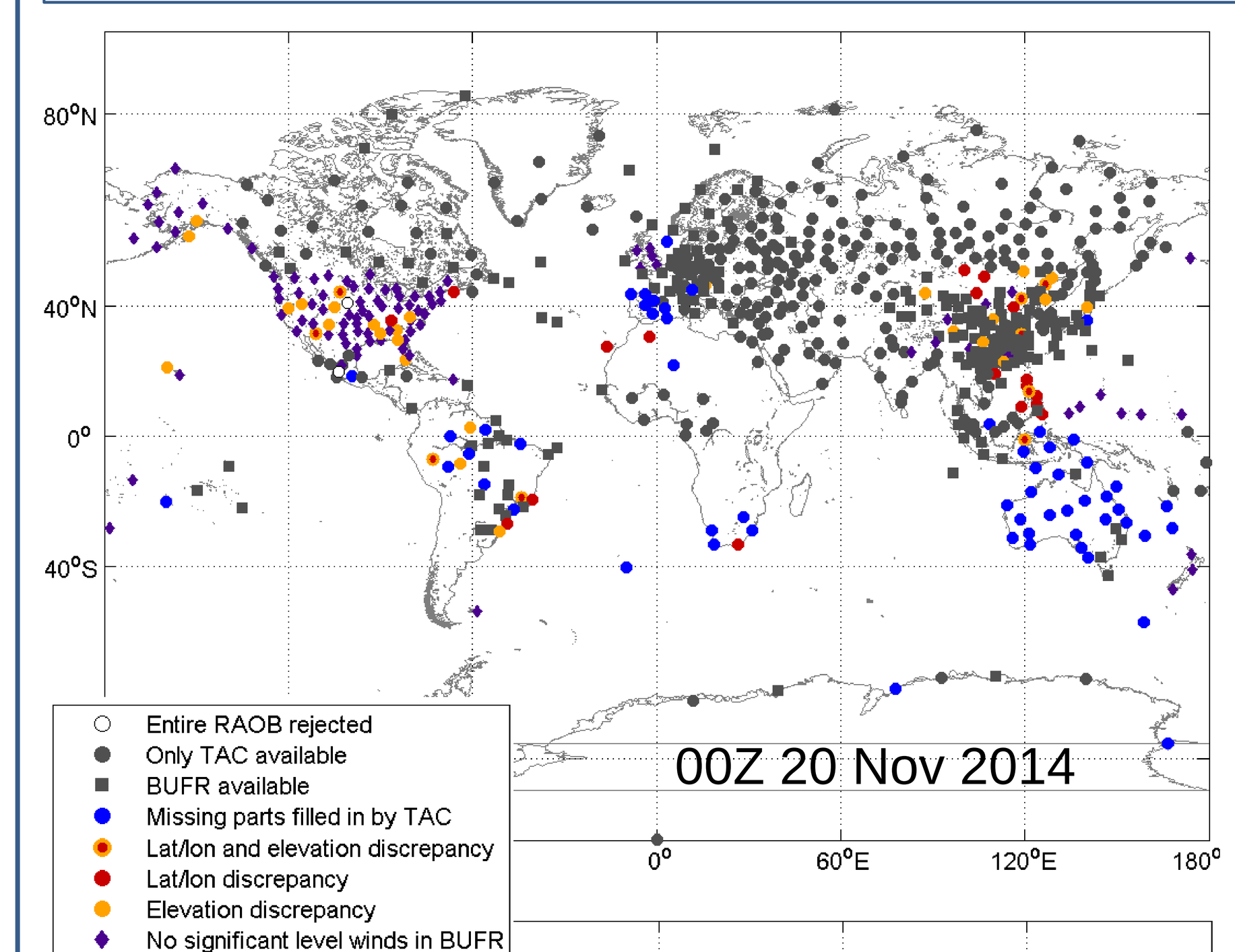
- The final BUFR message from a given station should include the whole sounding (purple), according to WMO rules.
- Some countries put each TEMP part in a separate message—"BUFR by parts" (red). Many NWP centers will not use these.
- Other countries use separate messages for the sounding below and above 100 hPa (cyan).
- A few countries only encode data below 100 hPa (green).

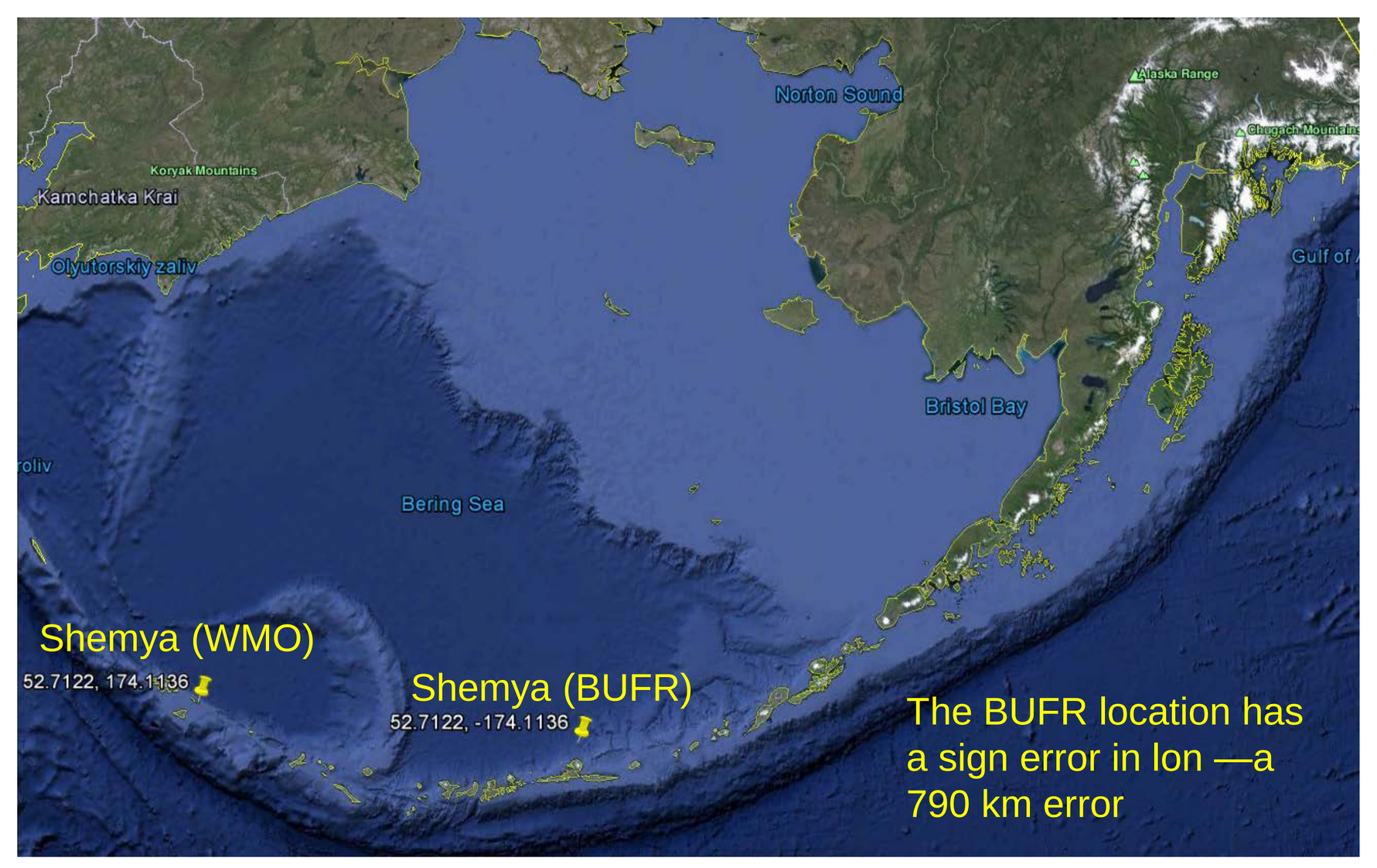


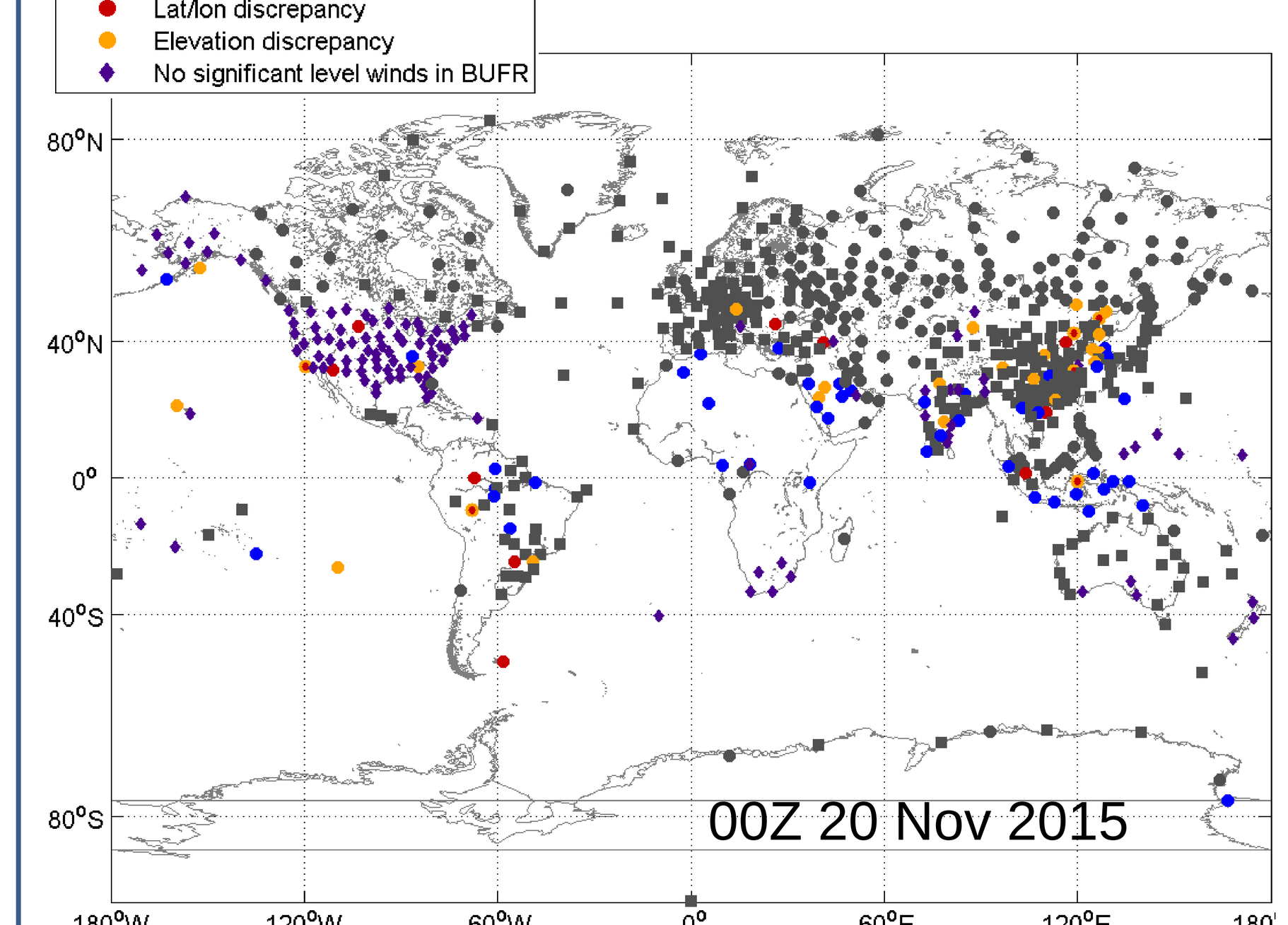


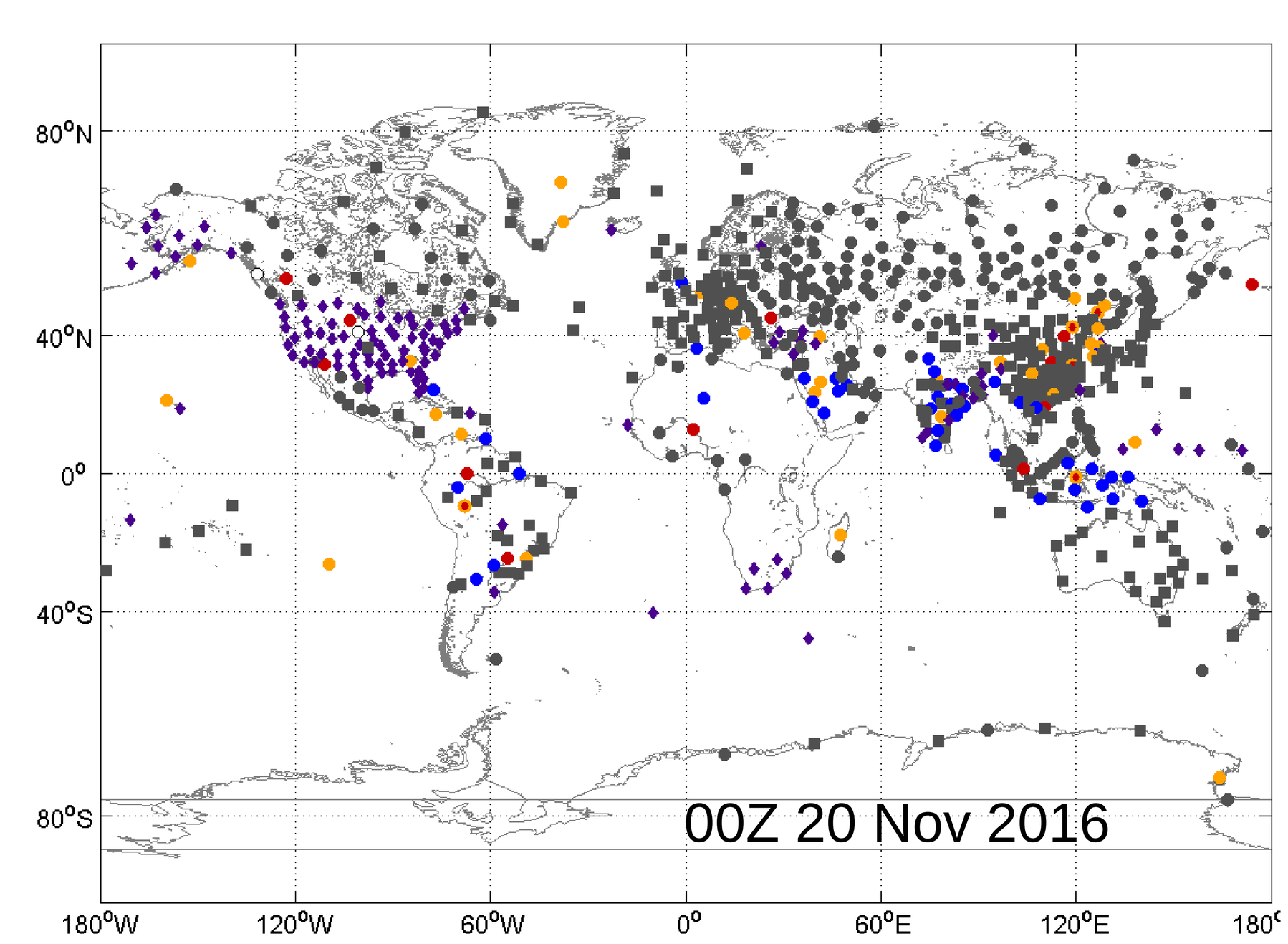
BUFR data issues

- Errors are present in the BUFR lat/lon (red) or elevation (gold), primarily in reformatted BUFR where an external station list is needed.
- Some countries do not include the significant level winds with the rest of the sounding (purple).
- Incomplete BUFR soundings can be filled in with TAC data (blue).



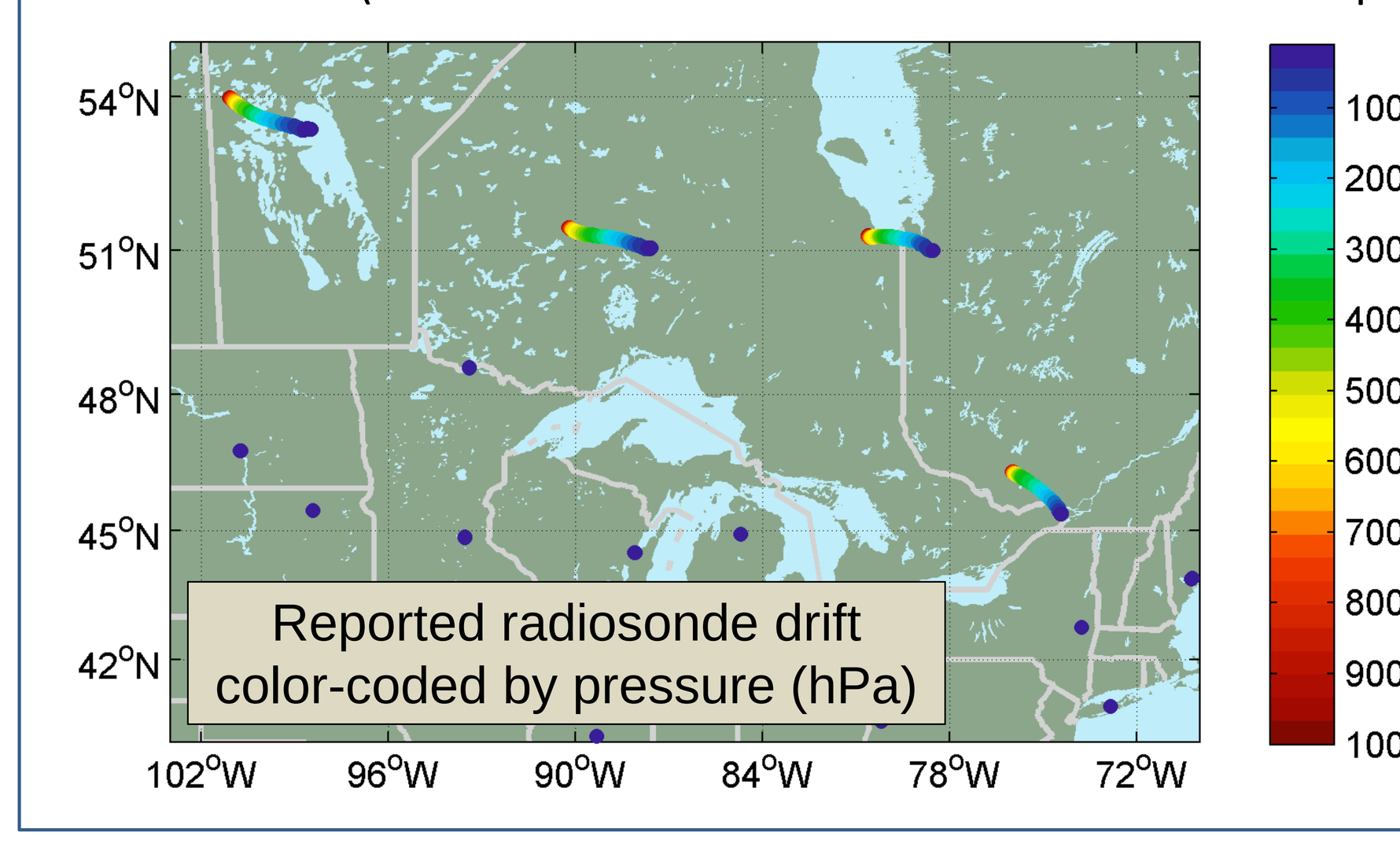


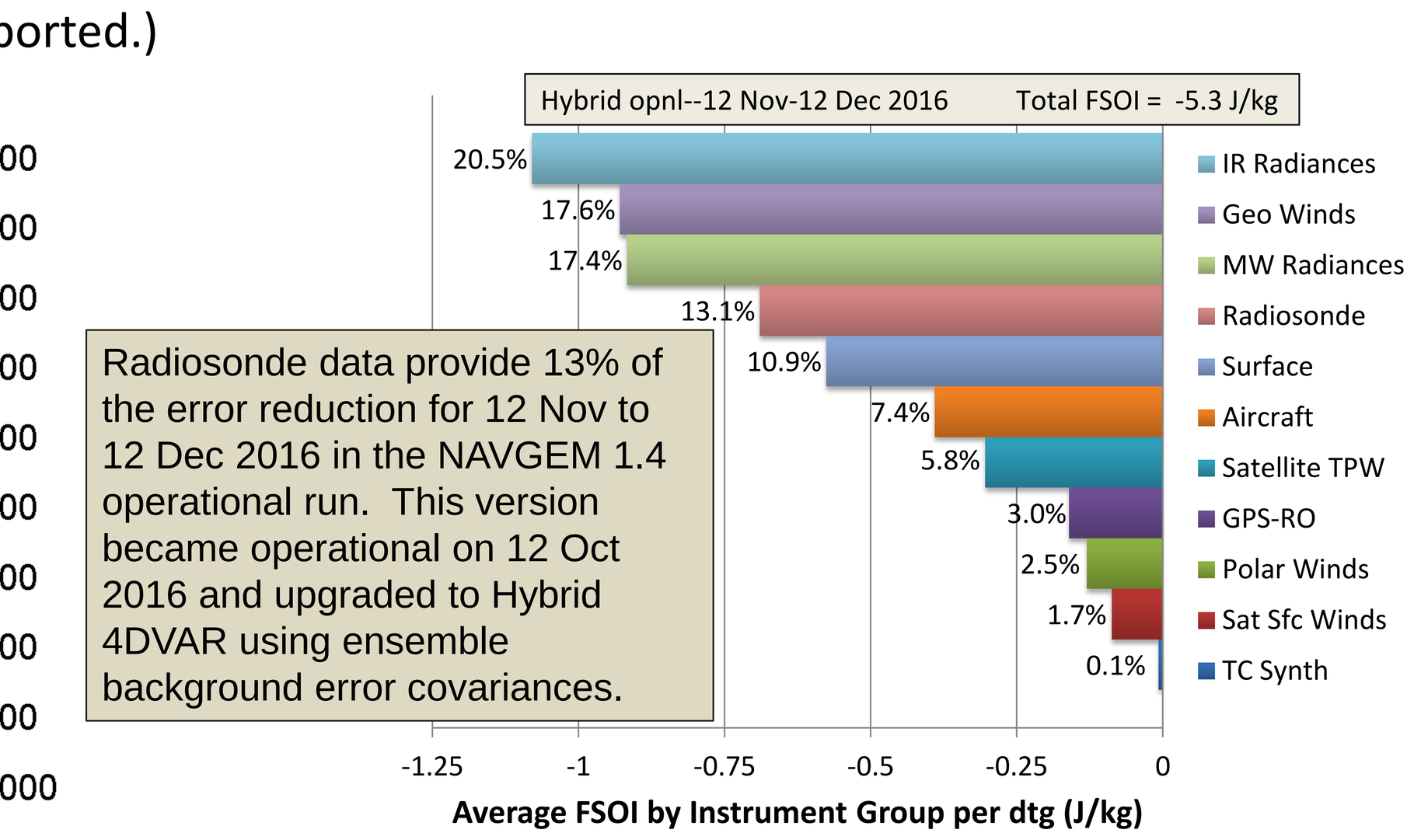


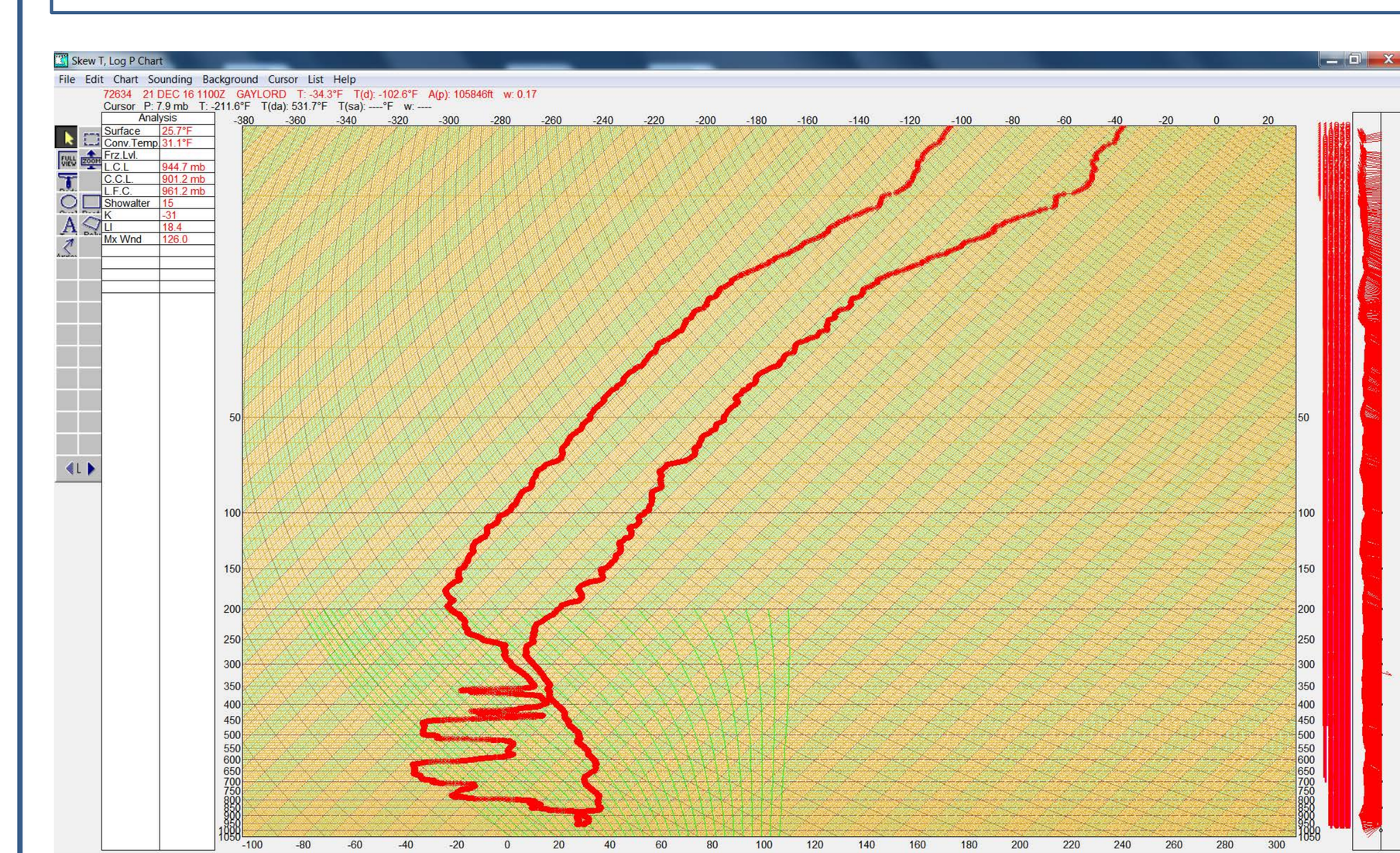


Use of BUFR radiosonde data in NAVGEM

- Decoded data are read and checked for errors in both data and metadata.
  - Data errors
    - Climatological/physical limits checks
    - Check for known errors
      - Pressure divided by 10 for significant level winds above 100 hPa from some stations in India
      - Significant level wind speeds divided by 2 from Viet Nam, etc.
  - Metadata errors (compared to FNMOC station list)
    - Errors converting deg-min-sec to decimal degrees (e.g., Yuma, AZ (32° 50′ converted to 30.50° instead of 30.86°)
    - Sign errors in longitude (e.g., Shemya, AK, is degrees W instead of E)
    - Use of Hp (barometer height) vs. Ha (field elevation)
      - In the U.S., pressure at Hp is adjusted to the elevation Ha, so Ha should be used as the station elevation for U.S. stations even though Hp is usually used as the station elevation by NWP centers.
      - e.g., Kodiak, AK, has Hp = 33.8 m and Ha = 5.5 m—reformatted BUFR uses 34 m as the station height, but 6 m should be used since the surface pressure is adjusted to the radiosonde release point
    - Errors in station list used for reformatted BUFR (e.g., Rapid City, SD, located 60 km too far north)
- TAC and BUFR data are duplicate-checked separately, then duplicate-checked again after TAC and BUFR data are consolidated. All TAC data are rejected when native BUFR are available.
- The surface elevation is estimated hydrostatically to aid in metadata reconciliation when BUFR values differ from the values in the FNMOC station list.
- High-resolution data are thinned to approximately 200 levels per sounding.
- The combined dataset is then subjected to quality control and assimilated in NAVGEM in the same way TAC data were processed previously, using NCEP's Complex Quality Control.
- BUFR radiosonde data have been used operationally since 23 Sept 2015 in NAVGEM's 4DVAR (NAVDAS-AR).
- Adding BUFR data increased the ob count (counting T, q, u, and v as separate obs) by ~30k at 00Z and 12Z.
- Incorporation of balloon drift is planned for NAVGEM.
  - Balloon drift will be used from reports where available and will be estimated from reported winds otherwise.
  - Example shows how reported drifts differ geographically for stations in southern Canada for data from 00Z 17 March 2015. (U.S. stations do not show drift since it is not reported.)







Observational Needs

As seen in the FSOI (Forecast System Observation Impact—Langland and Baker, 2004) figure above, radiosonde data are a critical observing system for NWP. While satellite data (radiances, satellite-derived winds, and GPS-RO) currently represent about 90% of the observations assimilated in NAVGEM, the previous figure shows that conventional observations (radiosonde, aircraft, and surface data) are responsible for 31.4% of the error reduction. NWP has an ongoing need for these operational conventional observations, not only in the U.S. but worldwide.

Future work with BUFR radiosonde data

High resolution real-time BUFR capability from NWS stations is currently being tested and should be deployed during spring to late summer 2017 (Aaron Poyer, Observation Systems Test Director, NWS).

- High-resolution data currently goes only to NCEI some hours post-launch using a legacy BUFR template.
- The NWS initiative will provide real-time native BUFR data using the standard WMO template and will include drift.
- Example shows data from a test launch from Gaylord, MI. Test sites include Pittsburgh, PA; Amarillo, TX; Gaylord, MI; Reno, NV; and Fairbanks, AK. Shreveport, LA, is scheduled to be added in January.
- 10 of the 92 NWS radiosonde sites are not included in this software upgrade, since they use a 403 MHz COTS system that was recently installed to mitigate RF interference with GOES-16.

References

Ingleby, B., P. Pauley, A. Kats, J. Ator, D. Keyser, A. Doerenbecher, E. Fucile, J. Hasegawa, E. Toyoda, T. Kleinert, W. Qu, J. St James, W. Tennant, and R. Weedon, 2016: Progress towards high-resolution, real-time radiosonde reports, **97(11)**. *Bulletin of the American Meteorological Society.* (http://dx.doi.org/10.1175/BAMS-D-15-00169.1)

ECMWF, 2016: TAC to BUFR Migration Wiki. (https://software.ecmwf.int/wiki/display/TCBUF/TAC+To+BUFR+Migration#app-switcher)