Technical Memo

ECMWF
European Centre for Medium-Range Weather Forecasts

926

# Diagnostics of radiosonde uncertainties

Bruce Ingleby

May 2025

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

*http://www.ecmwf.int/publications/*

Contact: library@ecmwf.int

## Abstract

Advanced data assimilation diagnostics have been used to estimate vertical uncertainty covariances for radiosonde reports, putting most emphasis on temperature but also looking at humidity and wind. In some cases the assumptions involved are not met and the vertical correlations come from another aspect of the assimilation system, such as forecast model biases. For Vaisala radiosondes the results suggest that the uncertainties are almost uncorrelated in the vertical – at least on the relatively coarse grid formed by the standard levels. For some other radiosondes these diagnostics have shown up problems with the reported profiles. The other strand to this work is to look at the uncertainties for reference (GRUAN) radiosonde profiles and to try to establish a link to the uncertainties of operational radiosonde profiles a) from the same stations and b) for other stations using the same type of radiosonde. Some effort is desirable to a) specify observation uncertainties for new radiosonde types and b) look at the possible benefits of explicitly modelling vertical uncertainty correlations for radiosondes.


NB. This document was largely written in 2018, it has been slightly revised and issued now as the prospect of submission to a peer-reviewed journal receded, due to the pressure of other work.


## Plain Language Summary

The uncertainties (sometimes loosely called errors) of radiosonde profiles need to be specified for data assimilation. They are used in quality control and assigning the weight to be given to each datum. Uncorrelated uncertainties are simpler to deal with than correlated uncertainties. Most of the data samples examined can be considered to have approximately uncorrelated uncertainties although there are some counter examples such as wind direction biases (dealt with by temporary rejection and feedback to the data providers). 'Desrozier diagnostics' are the main tool used to examine the correlations in this study. Uncertainty profiles for individual radiosonde ascents provided by the GRUAN network were also examined. For various reasons they do not currently seem worth using directly in data assimilation, although the average features provide useful guidance.


## 1. Introduction

To make the best use of any set of observations a good knowledge of the uncertainty of those observations is important; including the correlations between different levels and variables in a radiosonde context. In general radiosonde measurements are treated as having uncorrelated uncertainties but this has received very little study. The Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN) provides reference quality radiosonde data including uncertainty estimates. The GRUAN Vaisala RS92 product (Dirksen et al., 2014) identifies the most important sources of uncertainty and notes whether they are likely to be correlated or not partially addressing this issue. In the current study data assimilation diagnostics are used to provide information on the correlations of operational radiosonde data, but care is needed with interpretation of the results. As will be seen spatial representation uncertainty is a major factor in the radiosonde diagnostics. Estimates of radiosonde uncertainty correlations could be used to improve data assimilation, but another motivation is mapping the GRUAN uncertainties into radiance space to help with the calibration of satellite data (Carminati et al., 2016; also, part of the GAIA-CLIM project).

In data assimilation the 'innovation' y-$H(x^b)$ is central, providing new information from observations (the notation used here follows Ide et al., 1997). It is the difference between the observations y and the model background (short-range forecast) $x^b$ as interpolated to the observation location by operator $H$. For radiosondes $H$ is an interpolation in space and time (and in 4D-Var it involves the model propagation over the analysis time window). For satellite radiances $H$ includes radiative transfer calculations. There are uncertainties in each of y, $H$ and $x^b$ these are usually considered independent and described by covariance matrices E, F and B respectively. In practice the instrument-error covariance matrix E and the representation-error covariance F are usually combined into an observation-error covariance matrix R where R=E+F (Ide et al., 1997). The meteorological literature tends to use the term 'observation (or measurement) errors', here we use the metrological term 'uncertainties' – the 'truth' is never known exactly. For the forecast model the 'truth' is (in some sense) the best representation of reality achievable by the model, this depends on the resolution of the model (Lorenc, 1986; these issues are discussed in more detail in section 4). Less formally y-$H(x^b)$ is known as observation-minus-background or O-B, it is also useful to look at y-$H(x^a)$, observation-minus-analysis or O-A, where $x^a$ is the analysis. The square roots of the diagonal elements of R and B are known as $\sigma_o$ and $\sigma_b$ respectively. Much information on radiosondes and other observations can be found in WMO (2023) and the WMO Rolling Review of Requirements (https://space.oscar.wmo.int/observingrequirements ) includes target uncertainties for different variables and different applications. Practical aspects of radiosonde assimilation are reviewed in Pauley and Ingleby (2022).

Section 2 presents data assimilation statistics that provide an insight into radiosonde vertical uncertainty covariances – and other aspects of the data assimilation system. Section 3 looks at the measurement uncertainties for reference radiosondes (especially for temperature) and the link between these and operational radiosonde reports. Section 4 discusses representation uncertainty and section 5 presents a summary and possibilities for future work. An appendix briefly describes work to make temperature and humidity uncertainty a function of radiosonde type.

## 2. Diagnosed vertical uncertainty covariances

### 2.1. Data used

The choice of which period and which version of the assimilation system to use was complicated by various ongoing changes. Over recent years there has been a migration from alphanumeric radiosonde reports (TEMP format) to the binary BUFR format (Ingleby et al., 2016, 2018). BUFR allows for higher precision, more sampling in the vertical and the reporting of the position of each level. A radiosonde ascent takes about two hours and during that time it can move horizontally by 200 km or more. Only launch location is available for old-style TEMP data, so in data assimilation systems the profile has generally been treated as vertical and instantaneous. As the resolution and accuracy of data assimilation systems improve this approximation becomes less appropriate and better treatment of the drift can improve analysis and forecast performance (Laroche and Sarrazin, 2013). ECMWF processing of radiosonde drift (for stations where we are assimilating BUFR data) became operational in June 2018. For technical reasons the processing splits the profile into 'chunks' of 15 minutes each. The stratospheric O-B standard deviation statistics are improved significantly by between 5 and 10%, as hoped (Figure 1), wind is also improved by several per cent in the upper troposphere. The biases are also improved.

The radiosonde drift experiment used is for Nov 2016 – Feb 2017 (four months). Drift processing is applied only for the BUFR reports that report position at each level: mainly from Europe, Australia and New Zealand (more recently others including USA send such data). Vertical thinning is used to select the levels used in the assimilation (this particularly affects high-resolution reports; all available standard levels are amongst those selected for assimilation, but this wasn't the case for the first implementation of BUFR radiosonde assimilation). This experiment also had a revised radiosonde T $\sigma_o$ (as in 45r1, Ingleby et al., 2018) and weak constraint 4DVar for stratospheric temperature (Goddard et al., 2017, ECMWF newsletter). One experiment was run with the weak constraint switched off to check that this currently has little effect on the diagnostics (results not shown).

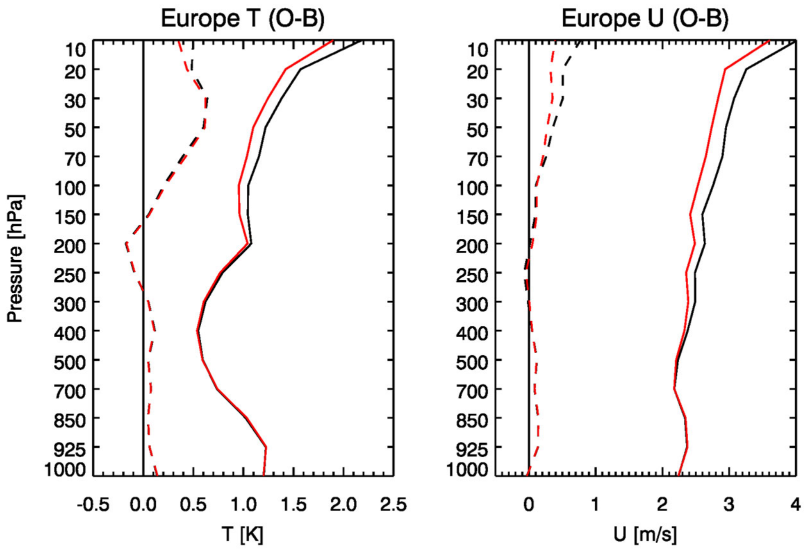



*Figure 1: Standard deviation (solid) and mean (dashed) radiosonde O-B statistics on standard levels for European stations with (red) and without (black) drift processing; temperature (K) and zonal wind (m/s), November 2016 – February 2017. Meridional wind has similar statistics to zonal wind and results for Australia plus New Zealand (not shown) were broadly similar but with slightly less improvement to the standard deviation.*

Figure 1 also provides a useful introduction to some of the climatological features of O-B statistics. Wind differences increase gradually with height, sometimes (not here) a local maximum can be seen around jet level. Temperature differences are relatively large in the boundary layer, have a minimum just above the mid-troposphere (~500-300 hPa in this case), often a maximum around the tropopause and relatively large values in the stratosphere, increasing quite sharply above 20 hPa. The upper tropospheric minimum in the standard deviation, ~0.5K in figure 1 (~0.4K if the statistics are computed just for German radiosondes), suggests observation uncertainties of 0.3K or less at these levels. If the measurement uncertainty is relatively independent of height in the free troposphere (see section 3) then the larger O-B differences above and below are due mainly to representativeness and/or background uncertainty. Further details of variations with latitude and radiosonde type can be seen in Ingleby (2017). In the extratropical lower stratosphere short range forecasts are biased cold – with a maximum of about 0.7 K at 50/30 hPa in figure 1. This temperature bias is mainly due to excessive water vapour and hence too much long wave cooling (Polichtchouk et al., 2021), the bias has approximately halved in the last few years. Krüger et al. (2024) use radiosondes to study errors in the ECMWF model round the tropopause.

It was decided to calculate statistics for the standard levels from 925 to 20 hPa (or 925 to 200 hPa for humidity), the levels at 1000 and 10 hPa were excluded because they would have greatly reduced the sample size (also when present 1000 hPa can be very affected by the proximity of the surface). When calculating vertical covariances matrices it is simplest if only profiles with all the levels are used – in this case the values need to be both present and assimilated, not rejected by the quality control. Particular attention is paid to Vaisala RS92 and RS41 radiosondes, the GRUAN stations have mainly used the Vaisala RS92 (Dirksen et al., 2014) and

RS41 (Jensen et al., 2016) radiosondes. In 2017 the RS92 was still the most widely used radiosonde (globally and in GRUAN), and both provided very good data (Ingleby, 2017) so their statistics will be combined.

### 2.2. Uncertainty covariance estimation

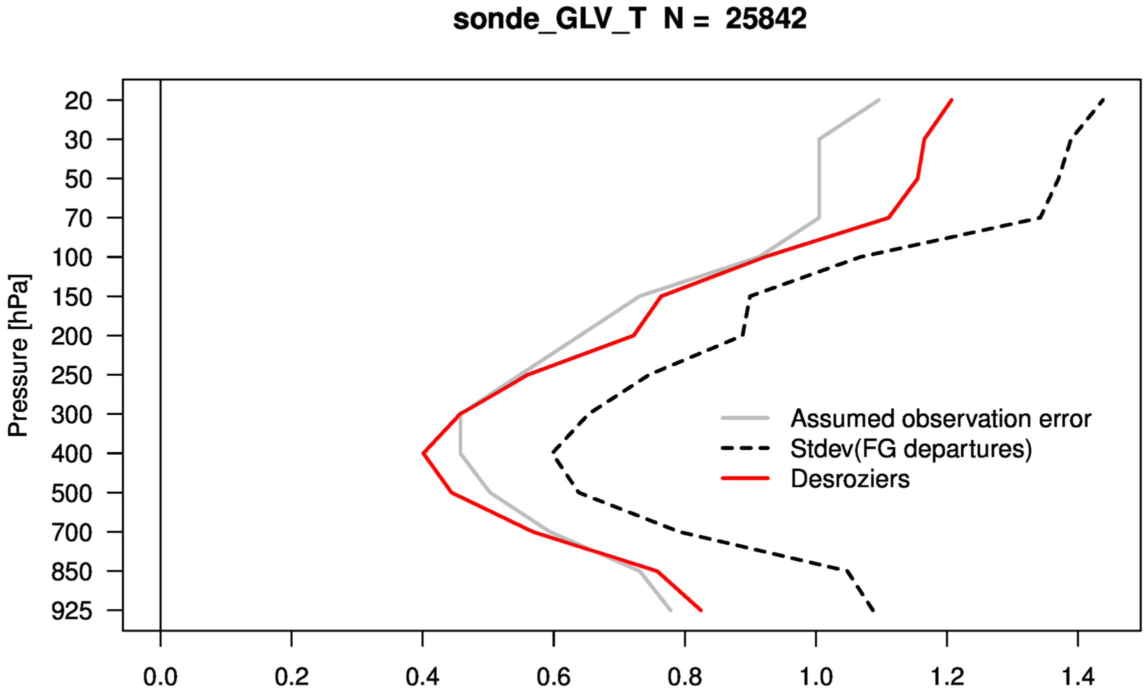
The statistics to be presented are based on the work of Desroziers et al. (2005) who used the covariances of observation minus background (O-B) with observation minus analysis (O-A) to improve the estimates of observation uncertainty covariance. There are various assumptions involved – notably that the observation weights are correct in the assimilation system. However, even when the assumptions are only approximately met it seems that the diagnostic can give useful information (Waller et al., 2016; there can be some problems with iterating the estimation procedure, updating the uncertainty estimates used in the assimilation, but that is not done here). At ECMWF the diagnostic has been used to estimate inter-channel uncertainty correlations for satellite sounding data, the correlations were then modelled in the assimilation system giving improved performance (e.g. Bormann and Bauer, 2010; Bormann et al., 2016). Similar work with satellite data been done at other centres, but the <(O-B)(O-A)> diagnostic does not appear to have been used for a detailed look at vertical correlations of radiosonde data before.

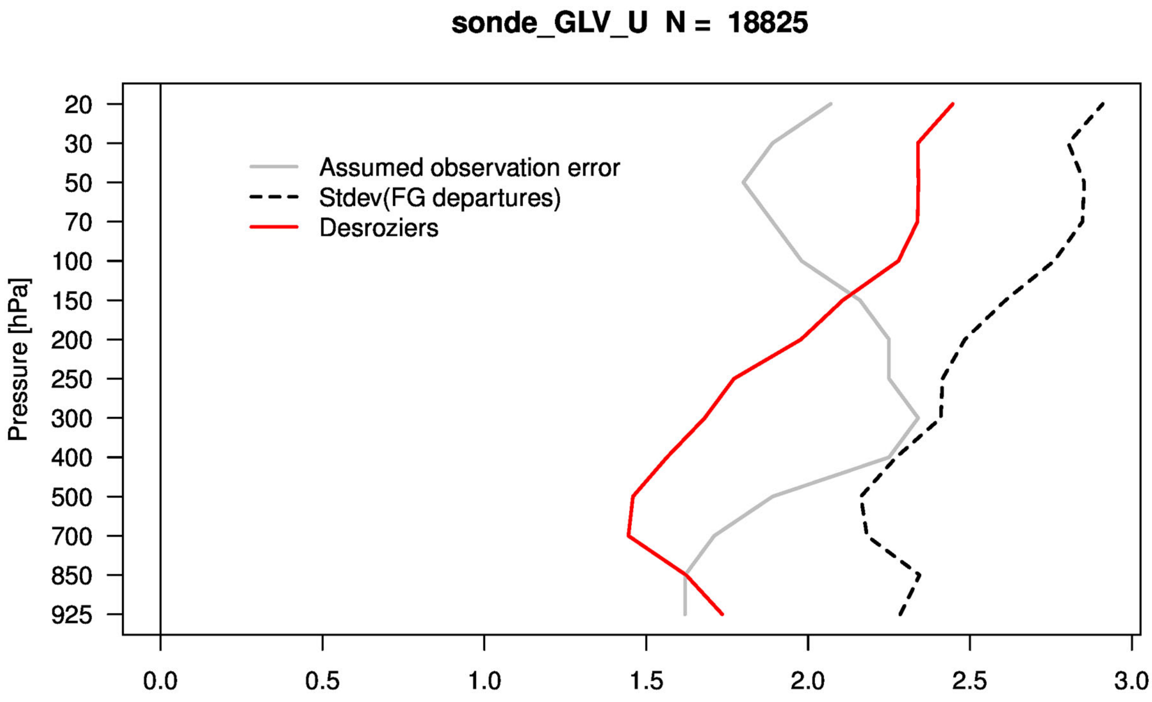
Hollingsworth and Lönnberg (1986) was very influential in promoting the study of $<(O\text{-}B)^2>$ covariances. Hollingsworth and Lönnberg (1989) is less well known but used $<(O\text{-}A)^2>$ covariances to assess the optimality of a data assimilation system. Both used radiosonde data, assumed that uncertainties from different radiosonde stations were uncorrelated and therefore looked at horizontal covariances including the extrapolation to zero distance from pairs of different radiosonde stations. In practice the Desroziers et al. (2005) <(O-B)(O-A)> covariances have largely been used to look at inter-channel ('vertical') covariances for satellite soundings. One practical issue is whether to subtract mean values before calculating <(O-B)(O-A)>, this is usual practice for satellite soundings which are subject to variational bias correction in the ECMWF assimilation system. On the other hand radiosondes are used as anchor observations and we often look at root-mean-square (rms) O-B statistics rather than the standard deviation (SD) of O-B. In this study most statistics are presented with the mean included, where the mean has been removed this is stated explicitly. (In the ECMWF system radiosonde temperature and humidity were bias corrected using night-time Vaisala RS92 – now RS41+RS92 - as a reference, and assuming that background biases are horizontally homogeneous, Agustí-Panareda et al. (2009); in practice the temperature corrections are generally small, Ingleby 2017).

If observation uncertainties are uncorrelated then this makes their representation easier and changes the information content of the measured profile. However if correlated uncertainties can be modelled well, then it is possible to take them into account and make the best use of the available information.

### 2.3. Estimated standard deviations for Vaisala radiosondes

The uncertainty standard deviations assumed and estimated for Vaisala radiosondes (with Vaisala processing, i.e. excluding the RS92-NGP used in the USA) are shown in Figure 2. This is for a set of standard pressure levels - only a subset of the assimilated levels. The grey profiles are the values assumed in the data assimilation, the dashed line gives SD(O-B) and the red line the profile estimated by the 'Desroziers method'. For temperature the input and estimated values are in very good agreement, except that above 100 hPa the assumed values are slightly too small. For zonal wind (and meridional wind, not shown) the $\sigma_o$ used has a maximum too low, at about 300 hPa and the values are somewhat low in the stratosphere. For specific humidity, q, the shape of the profile is in good agreement, but at low levels the assumed values are somewhat too high. Humidity is a difficult variable to specify uncertainties for: q varies by orders of magnitude in the vertical and also is much higher in the tropics than at high latitudes. The observation uncertainties are specified for relative humidity (RH) and then converted to specific humidity.



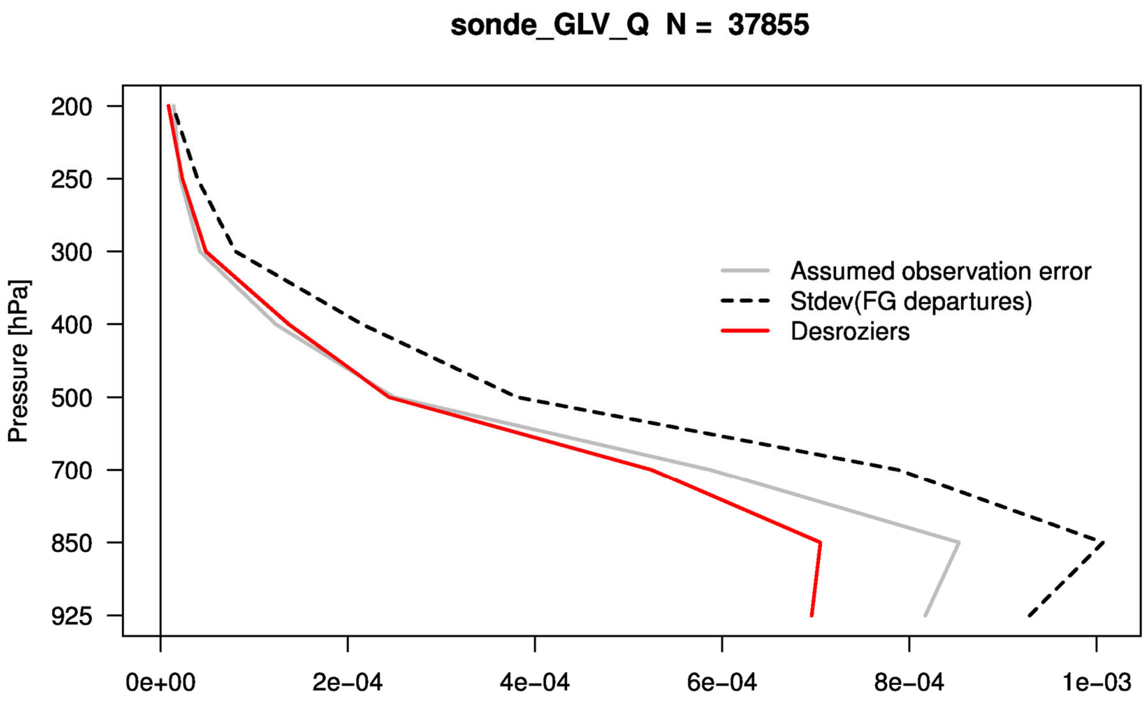



*Figure 2: Uncertainties for a) temperature (K), b) zonal wind (m/s) and c) specific humidity (kg/kg) for global Vaisala radiosonde standard levels, Nov 2016 – Feb 2017. The title gives the number of profiles used for the calculation (this is higher for humidity because of the more restricted set of levels used, the lower number for wind compared to temperature is presumably due to quality control rejections and possibly occasional gaps in the wind profiles).*

## 2.4. Estimated vertical correlations for Vaisala radiosondes

For simplicity wind vertical correlations are presented first. Figure 3 shows results for global Vaisala radiosondes, zonal wind component (results for the meridional wind component are very similar, not shown). To a first approximation the statistics for different levels are uncorrelated (at this vertical resolution). This is not too surprising: the retrieval of winds from GPS signals (as used by Vaisala and many other radiosondes) gives high resolution winds which are computed independently. (A case involving radar winds is examined in section 2.5.) Looking more closely there are small, positive (~0.1) correlations with adjacent levels in the upper troposphere (between about 500 and 250 hPa), there are even smaller negative correlations for some stratospheric levels. The RS92 winds (Vaisala processing) come from instantaneous measurements of the Doppler shift of the GPS signal, there is then some vertical filtering to remove pendulum motion from the winds (Dirksen et al., 2014). The vertical filtering is on shorter scales than the standard levels; if there were any vertical correlations in the measurements one would expect it to affect all levels more-or-less equally. If the correlations are not just noise, they presumably come from elsewhere in the assimilation system. Hollingsworth and Lönnberg (1989) showed that short-range positive correlations in $\langle(O-A)^2\rangle$ indicate that the assimilation system is underfitting the observations (assumed uncorrelated), the same should also be true of $\langle(O-B)(O-A)\rangle$. However, there are several possible reasons for this: $\sigma_o$ too large, $\sigma_b$ too small, spatial representation issues, overestimated length-scales of background uncertainty covariances or possibly ‘interference’ from another observation type (aircraft winds being the most likely in this case, although most are around 200 hPa so an effect at 500 hPa seems unlikely).

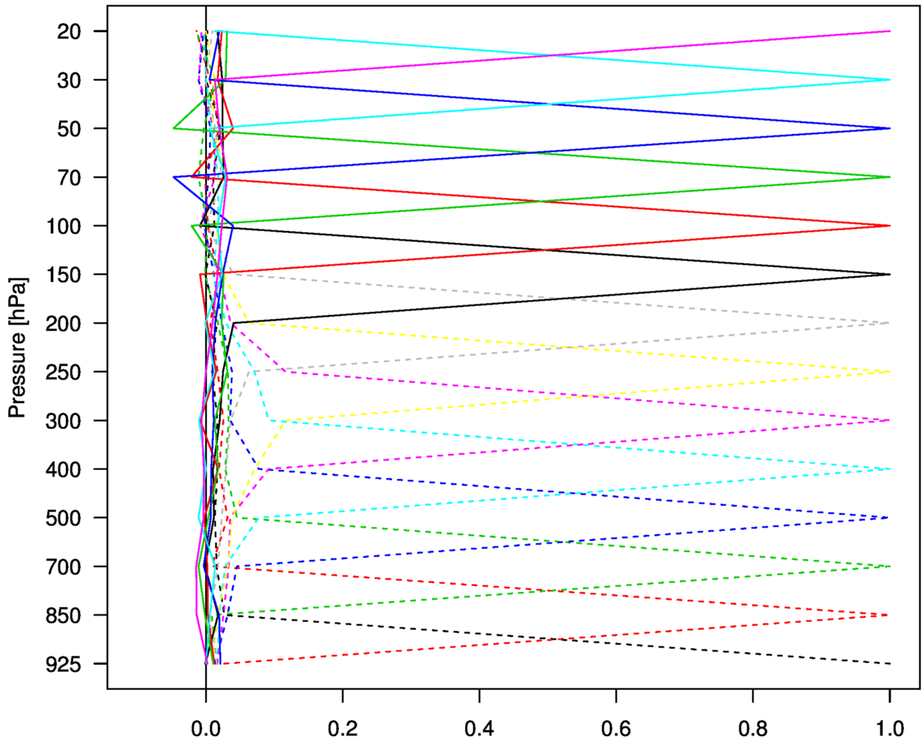



*Figure 3: Vertical correlations of zonal wind at radiosonde standard levels from 925 to 20 hPa for 'radiosonde drift' experiment November 2016 – February 2017 inclusive. For example, the black solid line shows correlations with 150 hPa. Vaisala RS92 and RS41 radiosondes included. Sample size N = 18825 profiles.*

Figure 4 shows vertical correlations of temperature for both global and tropical Vaisala radiosondes. The largest off-diagonal correlations are in the extra-tropical stratosphere, related to background biases there (discussed further below) – the analysis only moves part-way towards the radiosondes there. The tropical off-diagonal correlations are not centered on zero but are typically 0.05 to 0.2, the global mid-tropospheric values are about 0.05. The simplest explanation is that these are due to small calibration offsets for each radiosonde profile (see section 3). It is speculated that this is clearer in the tropics because of fewer inversions there and often the tropospheric temperature profiles lie close to the dry adiabatic lapse rate. The positive off-diagonal values drop to near zero at 200 hPa – almost certainly due to the assimilation of many aircraft temperatures at that level, the aircraft temperatures are typically biased slightly high even after bias correction. There is also a dip at 850 hPa, perhaps due to the top of the boundary layer in a large fraction of profiles.

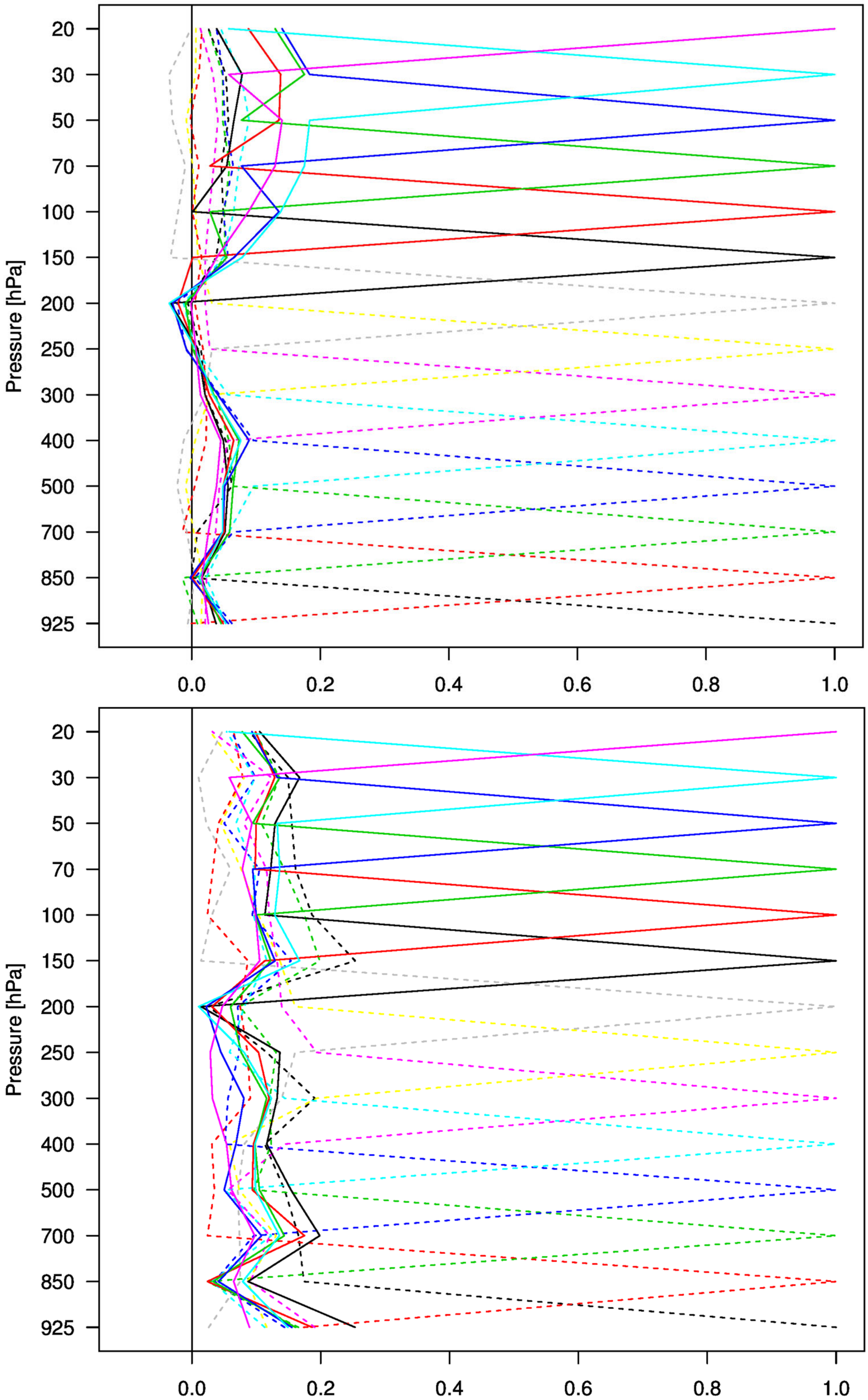



*Figure 4: As figure 3 but for Vaisala temperature reports. Top – global (N = 25842), bottom 20°N - 20°S (N = 3082).*

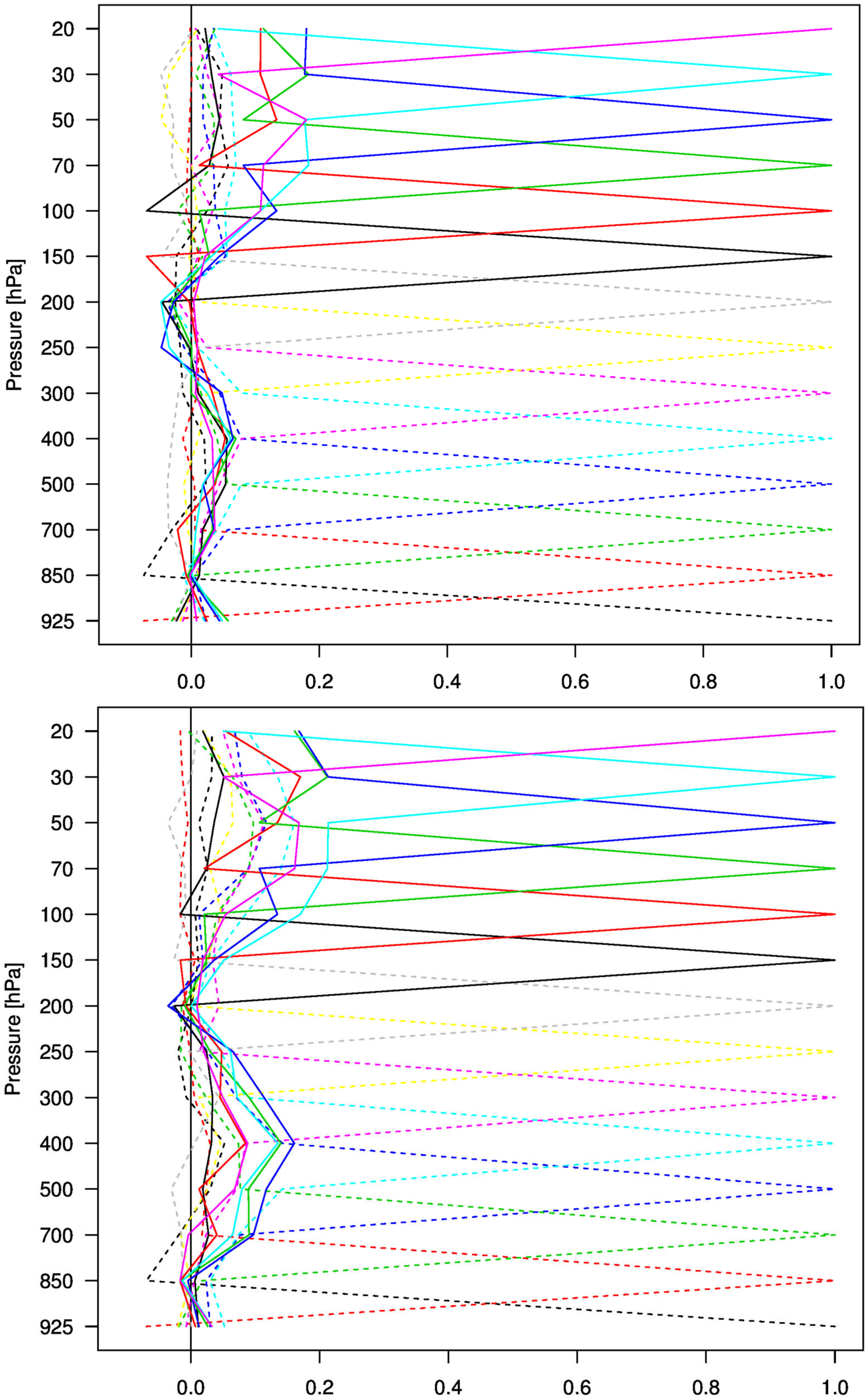



*Figure 5: As figure 4 except for European (latitude > 35°N, |longitude| < 40°) Vaisala radiosondes for 00(±3) UTC (N = 4032), top, and 12(±3) UTC (N = 4292), bottom.*

The diurnal cycle is investigated for European Vaisala radiosondes in figure 5. The positive off-diagonal correlations are larger in the daytime (12 UTC, bottom), relatively speaking this is more marked in the mid-troposphere than in the stratosphere which is slightly surprising as solar radiation effects are largest in the stratosphere (Dirksen et al., 2014). Figure 6 shows correlations from the same raw data (12 UTC; the 00 UTC results are similar, not shown) but with the means removed before the covariance calculation. This makes a large difference, the tropospheric off-diagonal correlations are near to zero (slightly positive for adjacent levels in the mid-troposphere), whereas stratospheric correlations are negative for adjacent levels suggesting slight over-fitting.

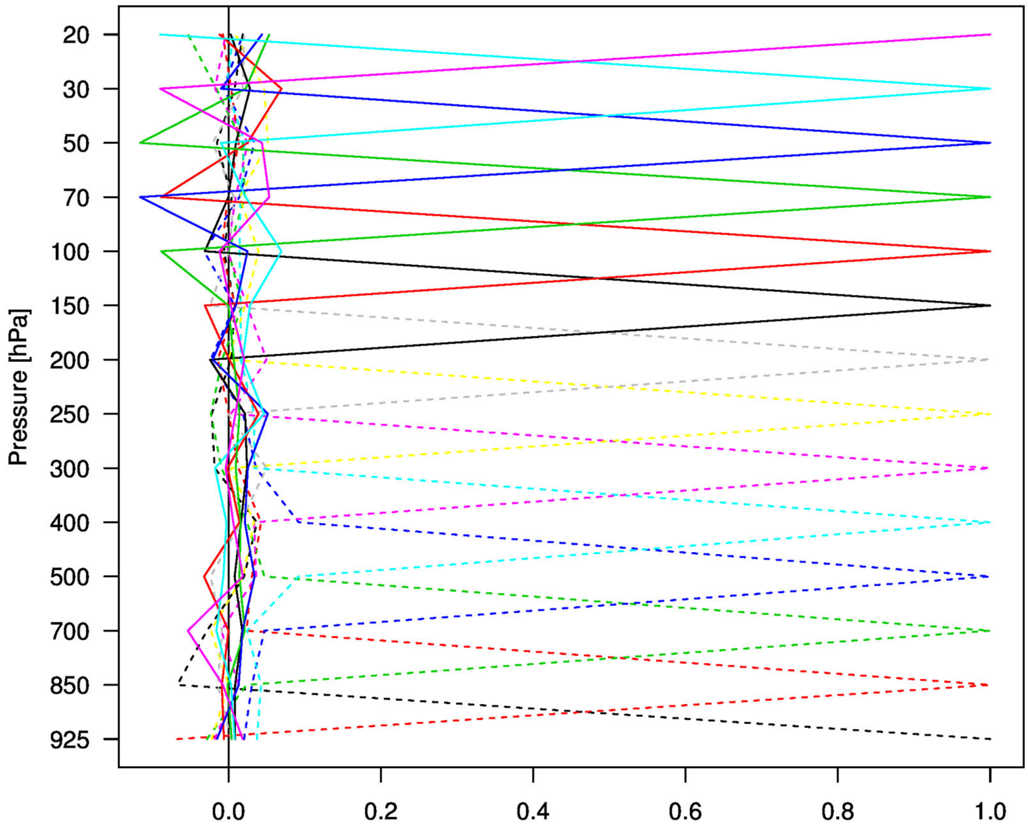



*Figure 6: As figure 5b, 12(±3) UTC, but with the mean removed before calculating covariances.*

Figure 7 shows correlations between temperature (T) and specific humidity (q) for Vaisala radiosondes. The most marked feature is the negative T-q correlation at the same level. This is down to about -0.3 globally at 850 and 700 hPa and below -0.3 in the tropics extending from 850 to 500 hPa. In the lower troposphere RS92 sensor time-lag is very small and the hydrophobic coating of the temperature sensor and heating of alternate humidity sensors minimises any wetting problems so this negative T-q correlation seems unlikely to be measurement uncertainty, and thus comes mainly from background and representation problems. Figure 8 shows example mid-latitude and tropical profiles with inversions that are misplaced or smoothed in the background – this tends to give errors of the opposite sign in temperature and humidity. Lorenc (2007) looked at similar issues in the Met Office system and discussed the difficulty of representing such features in the B matrix.

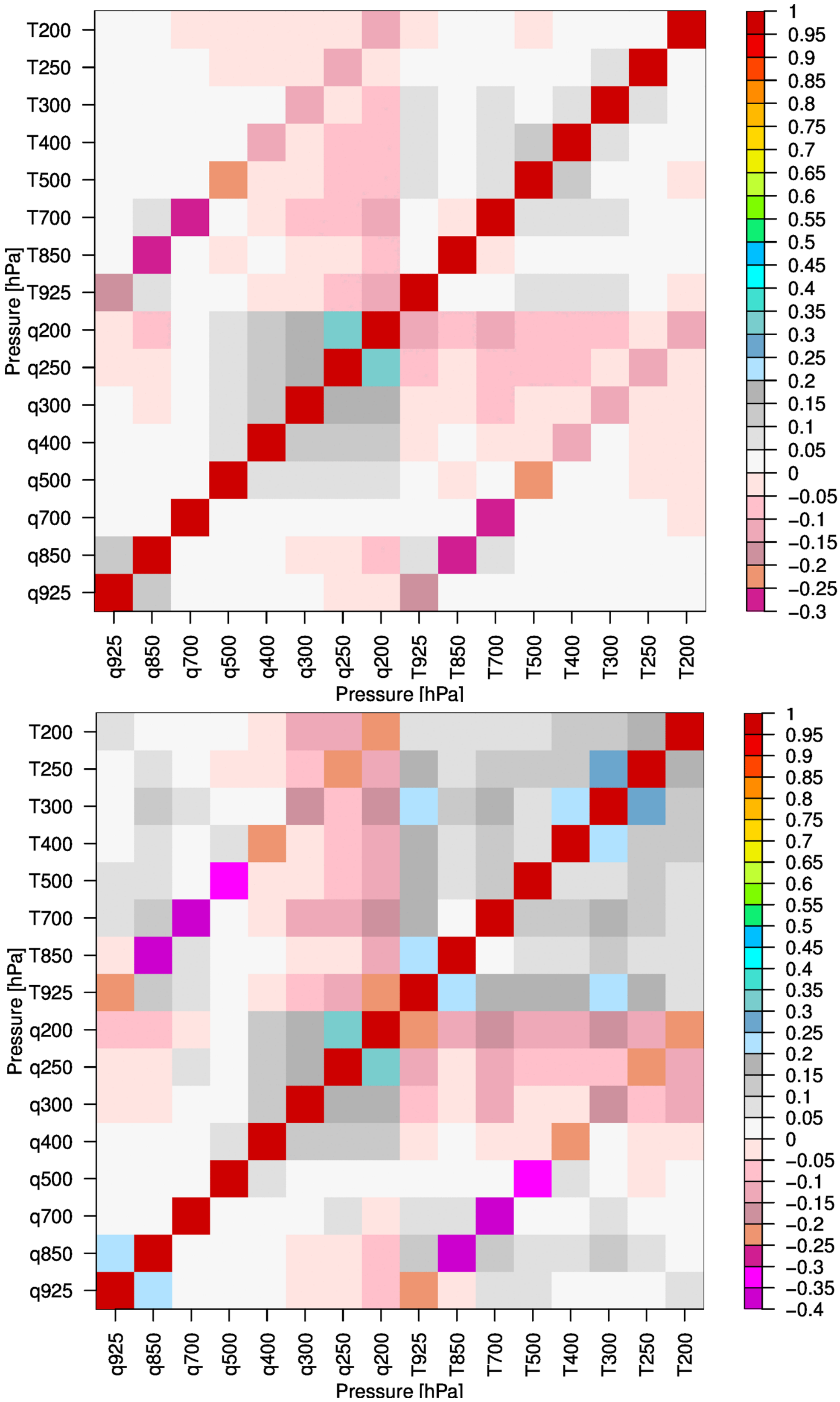



*Figure 7: Matrix plot of temperature and specific humidity correlations, 925 – 200 hPa. Top global Vaisala radiosondes (N = 36554), bottom Vaisala radiosondes 20°N-20°S (N = 6735).*

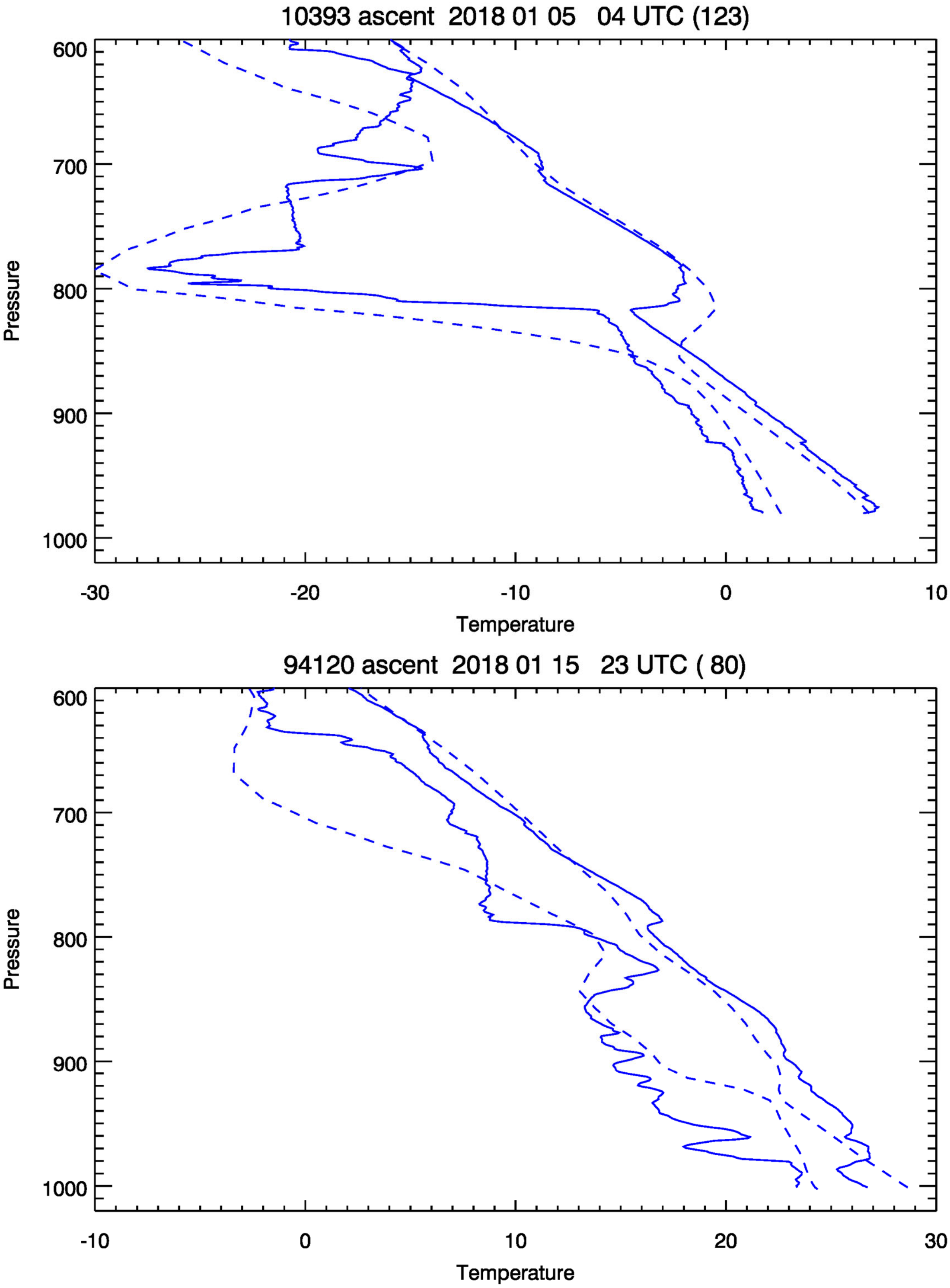



*Figure 8: Example profiles with inversions. Temperature and dew point temperature from high resolution reports (solid lines) and the ECMWF background (dashed lines). Top: 10939, Lindenberg, Germany, (52.22°N, 14.12°E) RS41; bottom: 94120, Cape Don, Australia, (12.42°S, 130.89°E).*

Figure 7 shows positive correlations for humidities between 500 and 200 hPa, reaching about 0.4 between 250 and 200 hPa. These upper tropospheric correlations are probably mainly observation related and due to the effects of solar radiation (Miloshevich et al., 2009). The small negative correlations between tropospheric temperatures and upper tropospheric humidity in Figure 7 are also probably due to diurnal cycle/radiation effects. In December 2010 Vaisala introduced a new version of the DigiCora software used with the RS92, this incorporated both a radiation bias correction and a time-lag correction for humidity (see the 'data continuity' pages are available via http://www.vaisala.com/en/meteorology/products/soundingsystemsandradiosondes/Pages/default.aspx). However it is up to individual meteorological services/stations as to if and when they implement new software versions and even in 2018 there is a mixture of the old and new processing versions in use (see section 3) and it is difficult to separate the reports. Figure 9 shows results just for Vaisala RS41 radiosondes, the RS41 includes a humidity time-lag correction and does not need a radiation correction because there is a thermistor within the humidity sensor. Jensen et al. (2016) showed that the main advantage of the RS41 compared to the RS92 was for upper tropospheric humidity. Figure 9 is consistent with this, the upper tropospheric humidity correlations are much smaller than in figure 7.

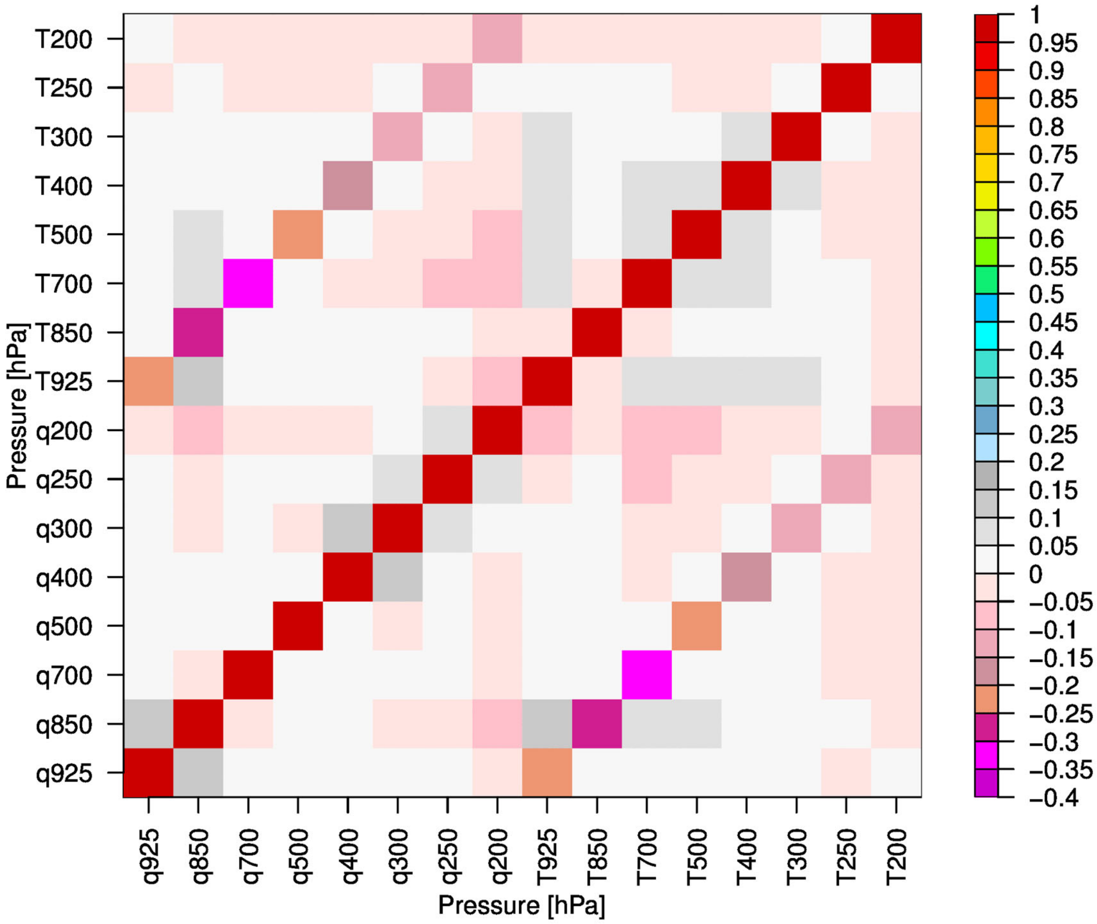



*Figure 9: As figure 7a but for global Vaisala RS41 reports (N = 11428).*

## 2.5. Chinese wind reports

When correlations were examined for Chinese radiosondes the most noteworthy results were for wind. Figure 10 shows the correlations for the meridional wind – the off-diagonal values reach about 0.35 in the upper troposphere (for zonal wind, not shown, the off-diagonal correlations reach 0.2). Suspecting that the problem was due to specific stations statistics were calculated for various sub-areas. For 25-30°N and 110-115°E the off-diagonal correlations for meridional wind reached 0.6. This area only contains five radiosonde stations, one of them is 57972 which appears to have a direction bias of about 10°, more-or-less constant in the vertical - Figure 11. Figure 12 shows mean reported and background winds at 300 hPa over China, for some stations they are almost identical so that the black arrow cannot be seen. 57972 (25.74°N, 112.97°E, directly above the scale arrow at bottom) has the largest direction difference, but three other stations (54292, 54342 and 59431) also have direction differences over 5°, these four are shown as red arrows. Results at other upper tropospheric standard levels are similar. The mean Westerly flow explains why directional errors produce a larger signal in the meridional wind than the zonal winds. For 57972 a direction difference from adjacent stations can be seen and it seems very likely that there is a radar orientation error – all Chinese radiosonde stations use radar for wind and position finding. Both 54292 and 54342 are near 42°N and their consistency with each other might indicate an error in the background winds. However, the direction differences are fairly constant with height (not shown) which is most consistent with observation direction bias.

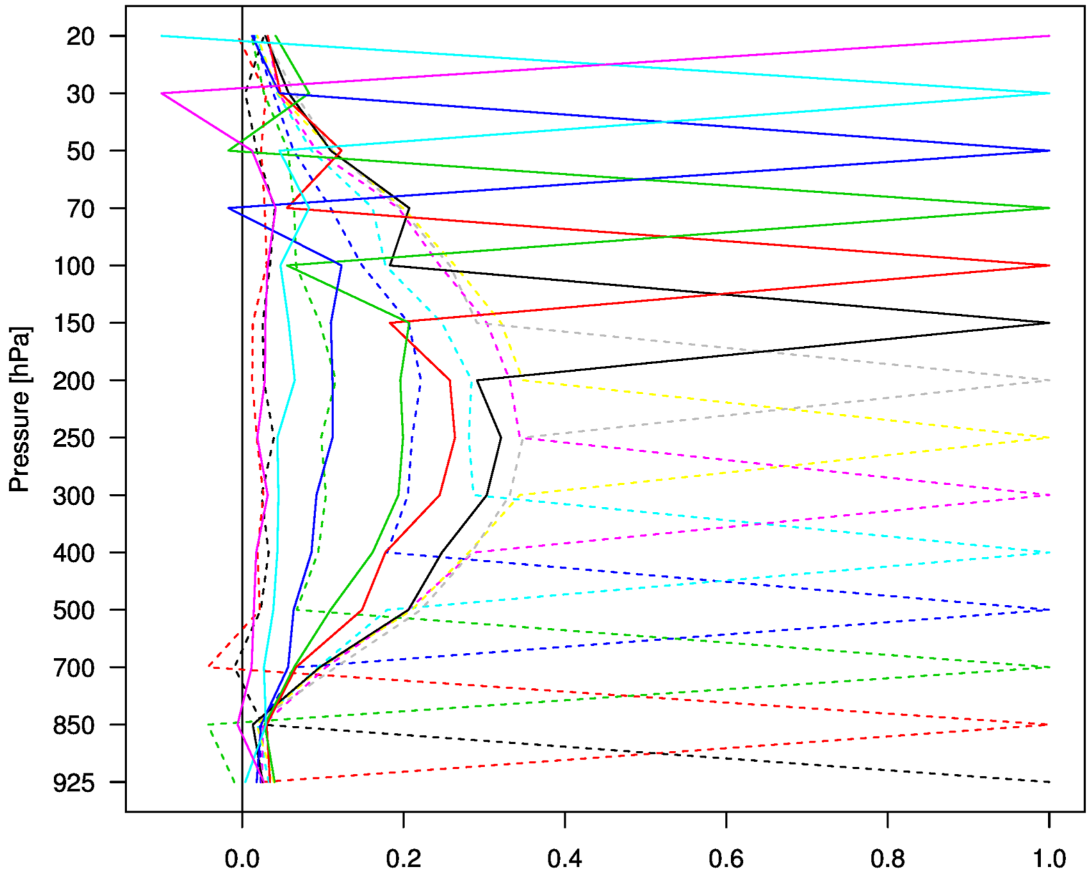



*Figure 10: Vertical correlations of meridional wind for Chinese radiosonde stations, other details as Figure 3. Sample size N = 9959 profiles.*

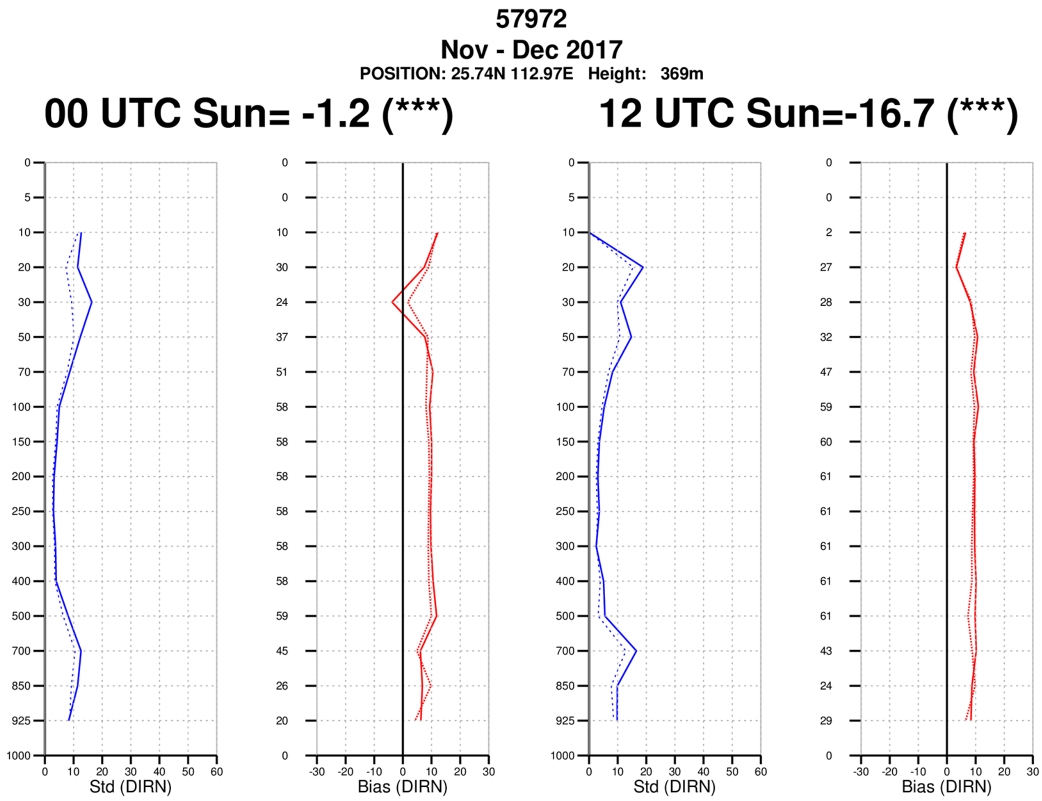



*Figure 11: Standard deviation and bias of wind direction (degrees) from 57972 relative to ECMWF background (solid line) and analysis (dotted line) for November – December 2017 (courtesy of Ersagun Kuscu). Statistics for 00 UTC (left) and 12 UTC (right), the numbers of reports at each level are provided to the left of the bias plots.*

Hollingsworth et al. (1986) provided the first documented examples of a data assimilation system being used to find errors in observations. One of those examples was a wind direction error for a remote island station. The bias in that case was larger – but as assimilation/forecast systems improve we can refine the methods for detecting suspect features. Over time there has been a trend away from radar wind-finding to the use of navigational signals, initially LORAN-C and OMEGA and now GPS. The main users of radar now are China and Russia (Ingleby, 2017, section 5.1). (For part of June 2017 one Russian station, 21432, had a direction bias of ~50°, noted by Brad Ballish of NCEP.) What seems remarkable in this case is how strong the signal is in the vertical correlation statistics (Figure 10) when a few stations (out of almost 100 Chinese stations) have modest direction problems. The issue was communicated to the China Meteorological Administration and corrections were made. For good detection of biases in wind direction (or speed) more targeted diagnostics are needed that integrate over vertical levels (e.g. from 700 to 100 hPa), an example is shown in Ingleby 2022: wind direction biases are mainly seen (then and now) in south-east Asia and west Pacific islands. Many of the affected stations only report wind and have limited vertical coverage (plus generally weak winds). GPS winds generally have uncorrelated errors (Figure 3) and perform well. There are occasional gaps in profiles, probably due to poor reception of GPS signals.

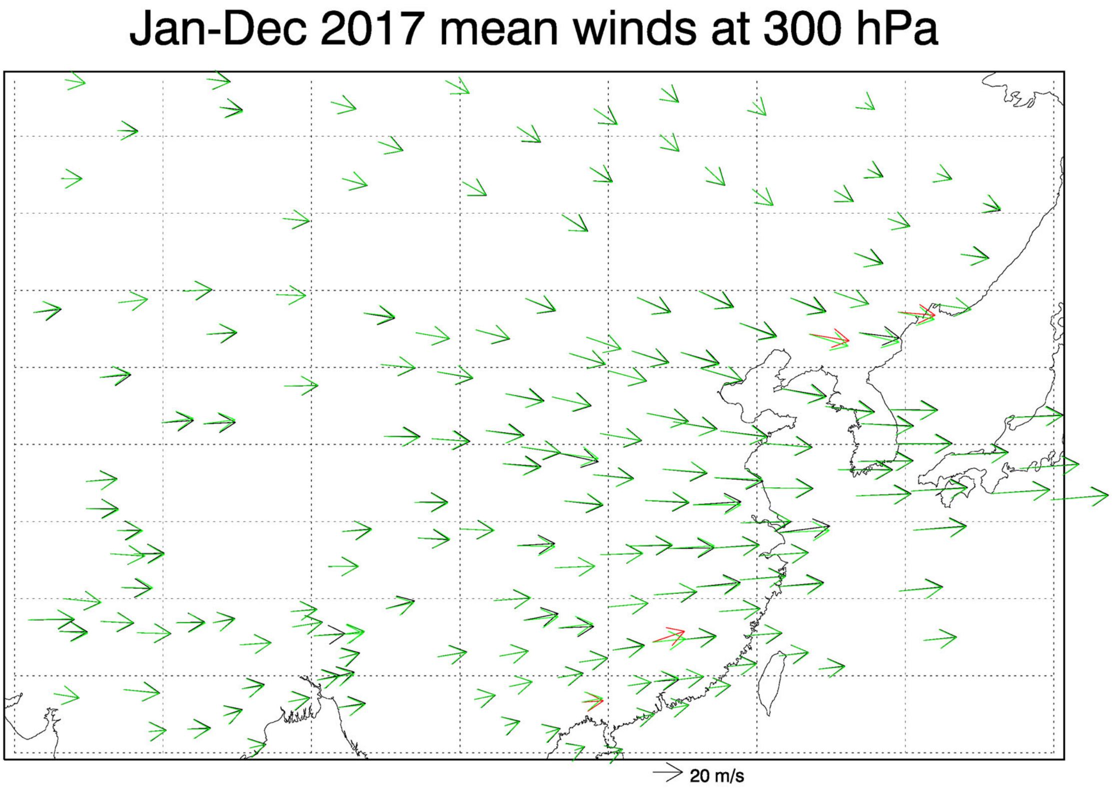



*Figure 12: Mean background (green) and reported (black, or red if the direction difference is more than 5°) winds for 300 hPa, 2017 data.*

## 2.6. Russian temperature reports

Figure 13 shows the computed vertical correlations for Russian temperature reports, these show quite large off-diagonal correlations and unlike the European Vaisala radiosondes (Figures 5 and 6) these are not 'cleaned up' by removing the mean values before calculating the covariance (not shown). Because Russian radiosondes use pressures derived from radar heights there may be problems at low radar elevation angles (Kats et al., 2005); however, this would probably give negative correlations between tropospheric and stratospheric temperatures – not seen. The simplest explanation is solar radiation effects (inadequately corrected), at short scales this may be exacerbated by the relatively slow response of the Russian thermistor. There are more than 10 different types of Russian radiosonde, they all had broadly similar temperature biases compared to the ECMWF background and their quality is worse than most other radiosonde types (Ingleby, 2017, section 3.5). An ECMWF study halving the use of Russian radiosondes (in response to a temporary cut in Russian soundings) found that they were beneficial, it appeared that low level temperatures in winter gave much of the impact (Ingleby et al., 2016). In 2019 the Russians started using a GNSS radiosonde (sonde_type=119) at some stations which has somewhat better quality than the older Russian types (not shown). The older types are still in use at many Russian stations. ECMWF tries to use BUFR reports, with extra levels, from the GNSS radiosondes but to use Alphanumeric reports, with fewer levels, from the older Russian types (this part of the data selection works by station identifier so will not always have the desired effect).

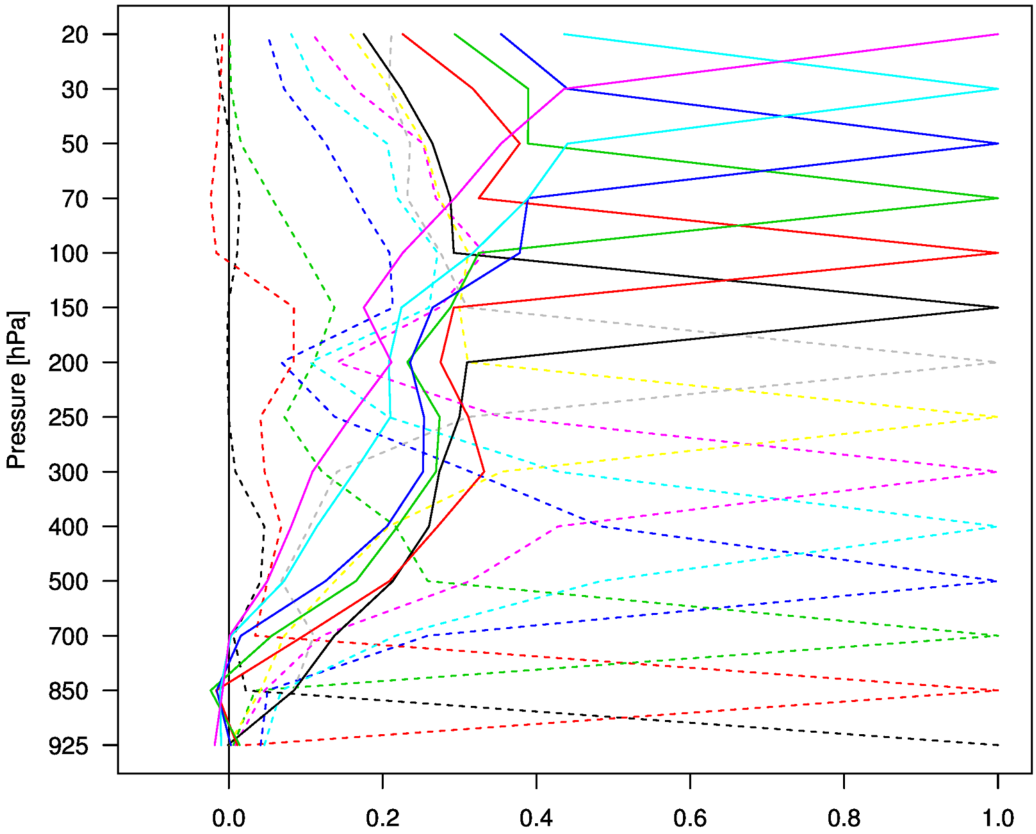



*Figure 13: As figure 3 but for Russian temperature reports (N = 9534).*

## 2.7. American LMS6 reports

In the 2010 WMO radiosonde intercomparison the American LMS6 radiosonde came a close second to the Vaisala RS92 in terms of overall performance (Nash et al., 2011), with the main difference being in terms of upper tropospheric humidity. The same is broadly true for operational radiosondes in 2015/2016 (Ingleby, 2017), except that the RS41 raises the bar a little higher. The LMS6 was used at stations in the USA and various Pacific Islands, it was retired in 2022. Over the USA there are very high densities of aircraft reports and these may affect the results to some extent. The zonal wind correlations (Figure 14) are similar to those for the Vaisala radiosondes (Figure 3) but have somewhat larger off-diagonal correlations in the troposphere – perhaps due to aircraft data. The smaller sample size would imply larger error bars on the estimated correlations. Figure 15 shows tropospheric temperature and humidity correlations – generally slightly higher than the Vaisala correlations between the same levels (Figure 7a). Temperature at 200 hPa has negative correlations (circa -0.1) with most other levels including the stratosphere (not shown) – this is thought to be due to the effect of cruise level aircraft data (seen in the Vaisala results to a lesser degree).

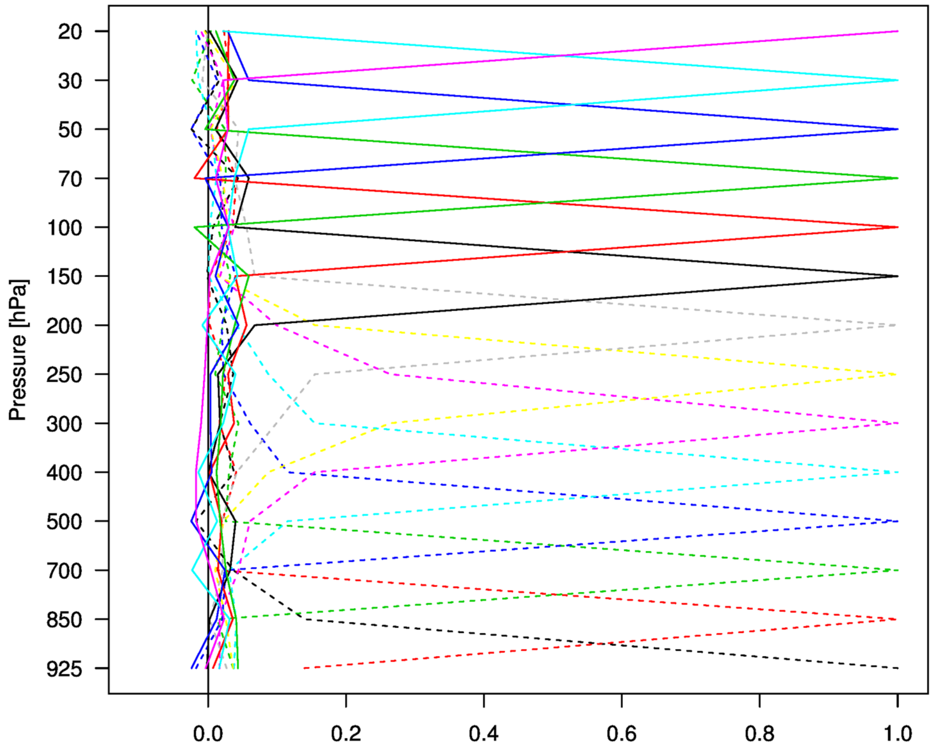



*Figure 14: As figure 3, zonal wind, but for LMS6 radiosondes (N = 4988).*

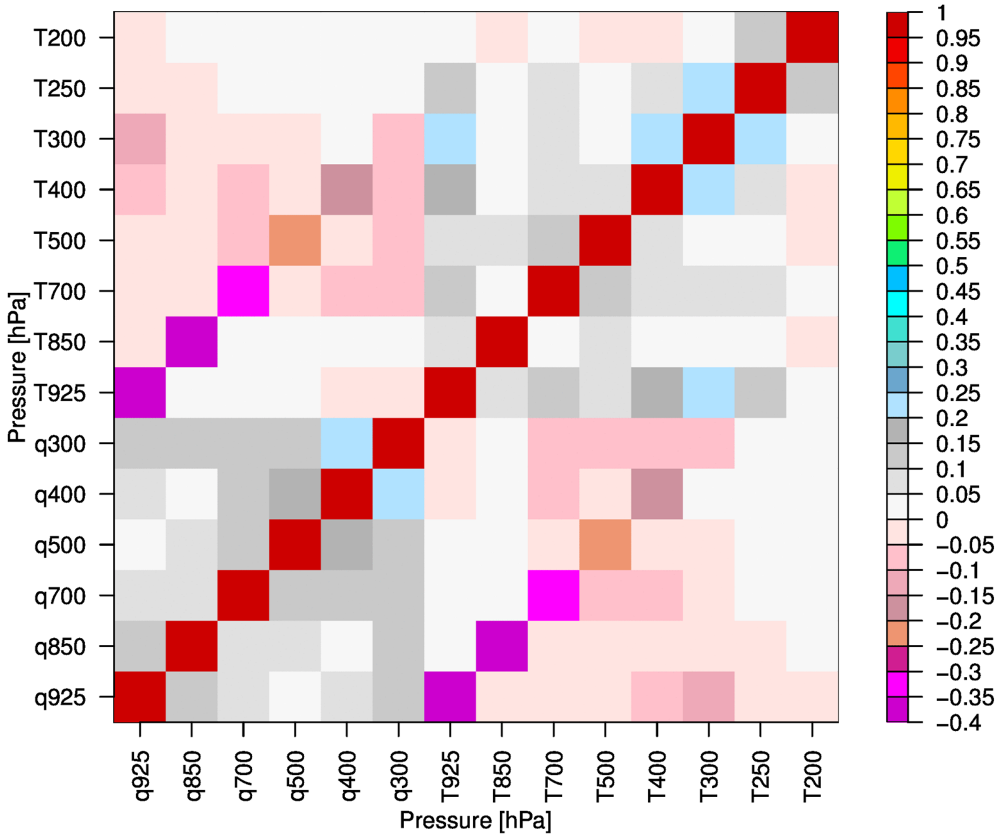



*Figure 15: As figure 7a but for LMS6 radiosondes (N = 4705). Note that humidity values at 250 and 200 hPa are not shown because they are not assimilated.*

### 2.8. High-resolution Vaisala reports over Europe

European stations were among the first to start high-resolution BUFR reporting and this allowed the production of diagnostics with higher vertical resolution; used pressures were rounded to the nearest 20 hPa and records from the same ascent with the same rounded pressure were averaged (pressures greater than 950 hPa or less than 10 hPa were omitted). The results can be seen in Figure 16 (temperature) and 17 (zonal wind). The results are fairly similar to the standard level results (Figures 3 and 4a) but are better shown using 'matrix' plots. There are fairly large correlations for the first off-diagonal but near-zero values at larger vertical differences: truer for wind than temperature. Positive correlations at short distance are probably due to correlated representation (interpolation) errors. If representation errors are uncorrelated then I think that (for optimal weights) <(O-B)(O-A)> will be slightly negative at short range as for <(O-A)2>, Hollingsworth and Lönnberg (1989). So there are two different effects (stronger at different levels?), further work would be needed to understand the details better.

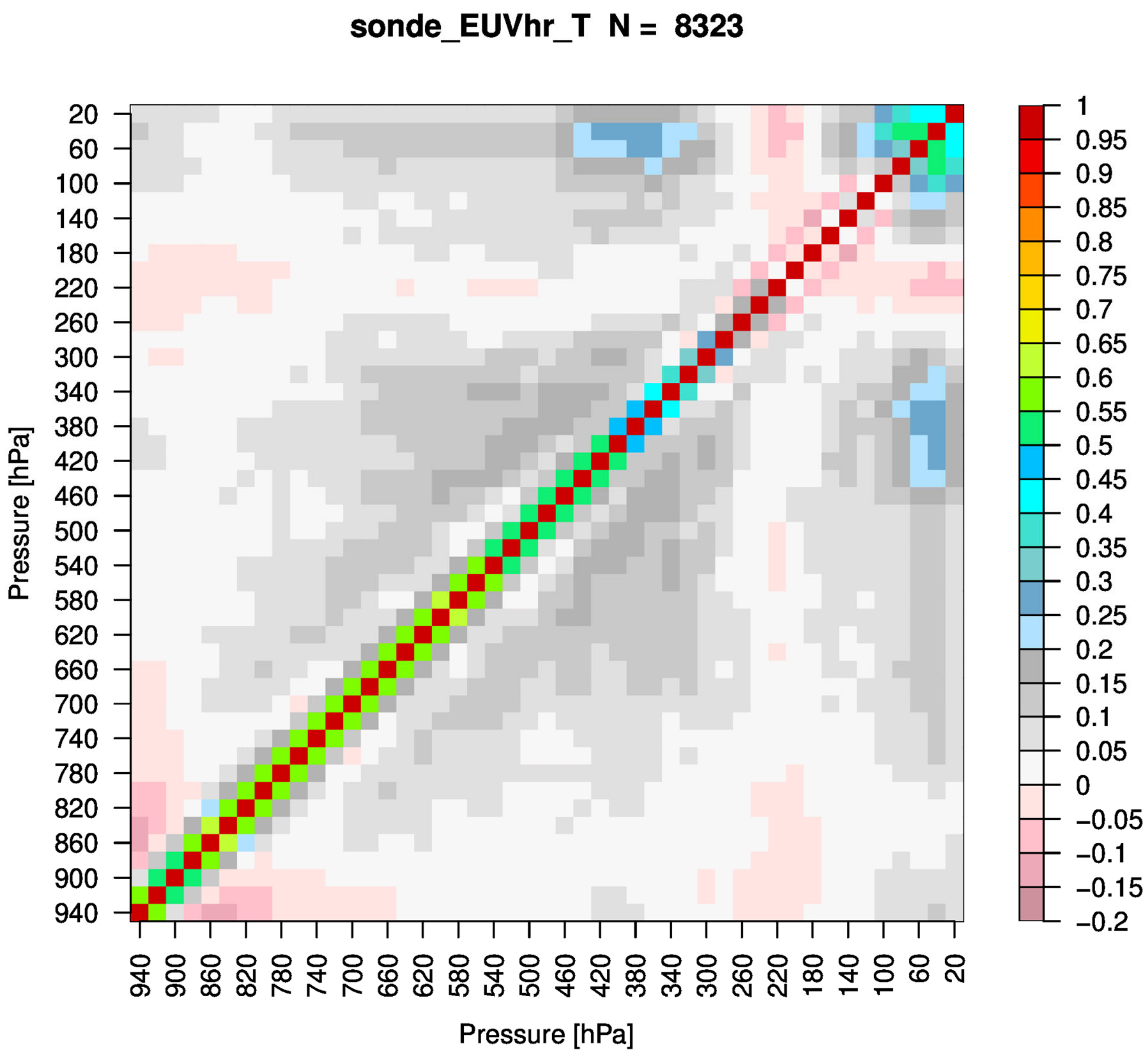



*Figure 16: Vertical correlations of temperature, 20 hPa pressure intervals, for Vaisala reports over Europe.*

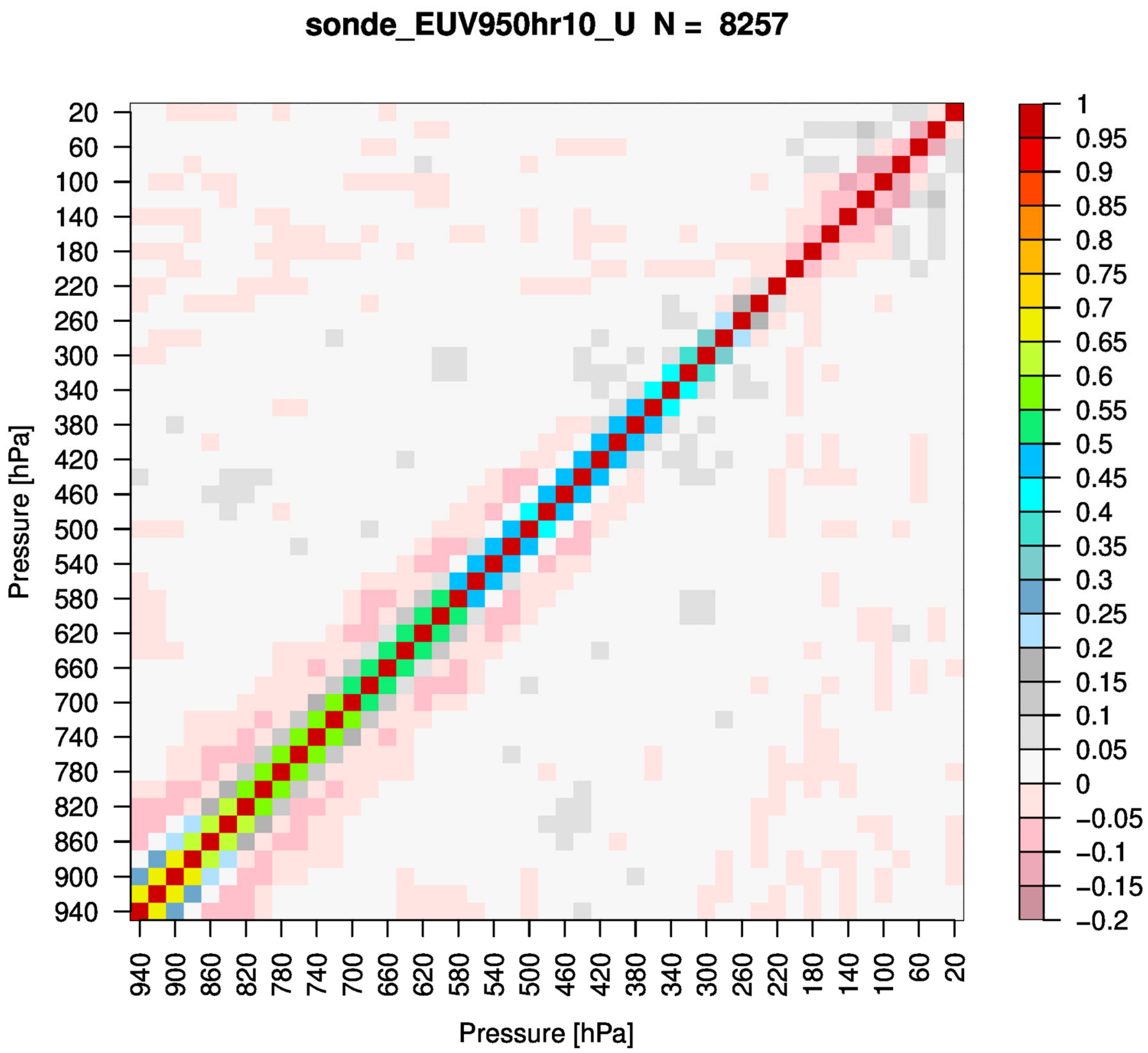



Figure 17: *Vertical correlations of zonal wind, 20 hPa pressure intervals, for Vaisala reports over Europe.*

# 3. Link between reference and operational radiosonde data

## 3.1. GRUAN uncertainties

Most GRUAN work up to 2018 was with the Vaisala RS92 radiosonde and Dirksen et al. (2014) describe the GRUAN RS92 data processing. The ambition is to have between 30 and 40 GRUAN stations globally, between 2013 and 2016 there were between about 10 and 15 stations – concentrated in the northern extratropics, unfortunately two sites in the tropical west Pacific closed in 2013/2014 (in 2019 Singapore became a GRUAN site). The lead centre is at Lindenberg, Germany (GRUAN abbreviation LIN), this reports every six hours. Another very active station is at Lamont, USA (also known as the Southern Great Plains site or SGP). One of the GAIA-CLIM activities has been to set up the 'GRUAN processor' – to look at the differences between GRUAN data and short-range forecasts (from both Met Office and ECMWF), and also to project the differences and the GRUAN temperature and humidity uncertainties into radiance space (Carminati et al., 2016). Metrologists have introduced the concept of a coverage factor (JGCM, 2008) and often use a factor of 2 (denoted by k=2). For a normal distribution $k$=2 will give a coverage probability of approximately 95 % (i.e. there is about 5% probability of the value falling outside the range specified). The k=2 uncertainties given by GRUAN should be divided by 2 to give standard deviations for comparison with the values used in data assimilation.

Figure 18 shows the diurnal cycle in the GRUAN temperature uncertainties at these two stations. As is well known there is greater uncertainty at upper levels during daylight hours – due to solar radiation effects and the problems of 'correcting' them. GRUAN has quantified this effect (for the RS92 radiosonde) and the uncertainty estimates seem to capture this quite well (Dirksen et al., 2014, Figure 10, as noted above they show k=2 uncertainty estimates). At low levels (but away from the surface) the main contributor to the uncertainty is the calibration. The other feature visible in figure 18 is the spikes (much larger in individual profiles) – coming from the temperature spike removal.

The low-level uncertainties for SGP (~0.14 K) are about 40% larger than those for LIN (these two fall near the extremes of the different GRUAN stations) which puzzled the author. The GRUAN lead centre provided an explanation: "The apparent discrepancy between LIN and SGP uncertainties can be explained by the different deviations of T found during the groundcheck. As described in Eq. (4) and the second row in Table 2 in Dirksen (2014), the combined uncertainty of the manufacturer calibration ( u_c(cal)=0.075 K, k=1) and the deviation found during groundcheck (Delta T_GC25 / 3) is designated as (absolute) calibration uncertainty (u_c,absolute (cal)). The factor of 3 is according to GUM. The groundcheck deviation is batch dependent, which explains the different uncertainties you found for 2016 of LIN and SGP below 15 km (as long as other sources are not relevant). However, contributions from the GC25 reference sensor also cannot be excluded. So, the groundcheck in LIN contributed a bit less to u_c,absolute (cal) than at SGP in this case. Note that for the RS92 sonde, the separate sensor boom is calibrated in the factory, whereas the groundcheck before launch is done with the complete sonde. This might explain why the average Delta T_GC25 is not zero." (von Rohden and Sommer, pers. comm., 2017). An offset in the reference thermometer used in the groundcheck at SGP could have caused the larger uncertainties, but it was not possible to confirm this happened. For non-GRUAN RS92 radiosonde reports it seems reasonable to assume that low level temperature uncertainties are at least as large as the larger GRUAN uncertainties, say 0.15 or 0.2 K (these are still quite small estimates).

The GRUAN uncertainties are for the measurement, without considering representation uncertainty. There is no explicit dependence on latitude or season, the main external influences are the solar elevation angle and pressure for temperature and humidity. The GRUAN RS92 uncertainty estimate for humidity is up to 6 % relative humidity and 0.4-1 m/s for wind speed, and 1 degree for wind direction.

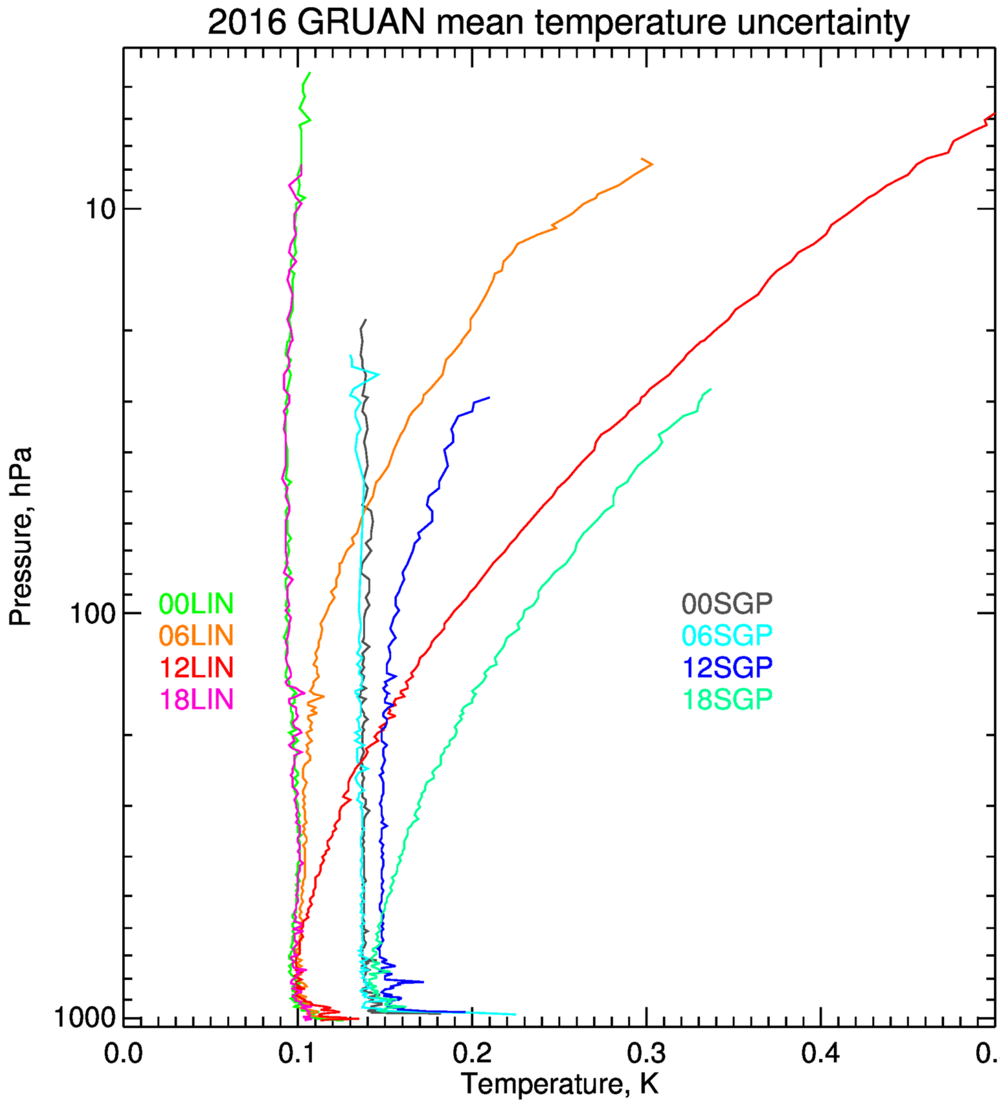



*Figure 18: Mean total temperature uncertainty (K, k=1) for Lindenberg (LIN) and Lamont (SGP) for 2016. Values calculated separately for 00, 06, 12 and 18 UTC ascents – see key.*

## 3.2. Differences between GRUAN and Vaisala processing

As discussed by Dirksen et al. (2014) the GRUAN and Vaisala RS92 processing give very similar results for temperature. The daytime radiation corrections differ but the net result is similar; GRUAN uses the Vaisala night-time radiation correction, which is small. The differences are larger for humidity largely because the GRUAN processing includes time-lag and radiation corrections for humidity. Pre-2010 versions of the Vaisala software did not include these humidity corrections. (Ironically GRUAN stations were still using older versions of the Vaisala software because the newer versions are not compatible with version 2 of the GRUAN processing). The corrections in DigiCORA 3.64 and later versions bring Vaisala processed humidities closer to GRUAN processed humidities (Yu et al., 2015), although there are still some differences. As already discussed, real-time high resolution BUFR reporting via the WMO Global Telecommunications System (GTS) is becoming more common. Figure 19 shows a comparison of the GRUAN profile (1 second data, black), the BUFR GTS profile (Vaisala processing, 2 second data, green) and the ECMWF background fields (red). For temperature the green line lies under the black line and cannot be seen, for dew point temperature minor differences can be seen. The ECMWF profile, whilst in good agreement, is obviously smoothed compared to the radiosonde profile.

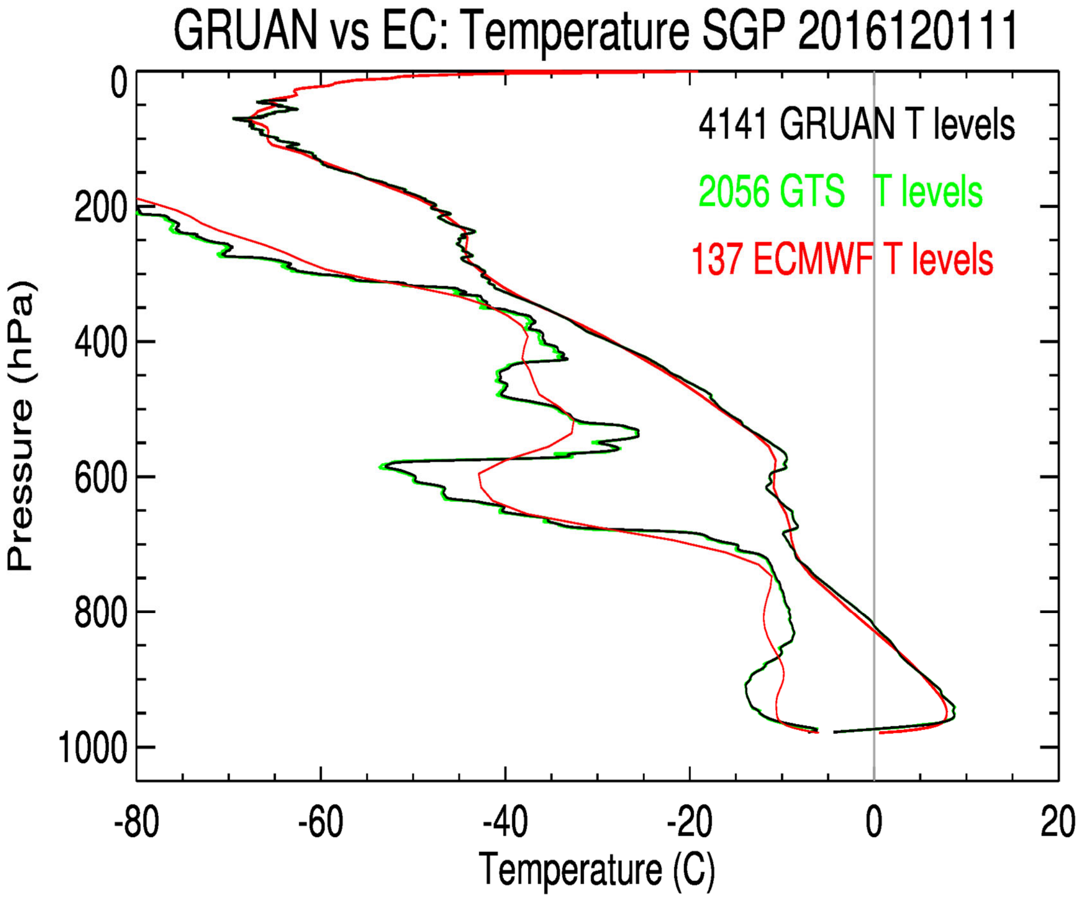



*Figure 19: Temperature and dew point temperature for the SGP (Lamont) station 11 UTC launch 2016-12-01. See text for further details.*

Some of the GRUAN stations have always provided real-time reports on the WMO GTS, others started following encouragement from GRUAN. For stations using the RS92 the real-time reports come from the Vaisala DigiCORA software, GRUAN versions from the same raw data are provided in delayed mode. Table 1 provides a partial summary of GRUAN stations (certified or candidate) reporting on the GTS in 2018. Since 2018 a few new GRUAN sites have been certified including Singapore and La Reunion.

| | Station | WMO id | Frequency | Country | Notes |
|---|---|---|---|---|---|
| BAR | Barrow | 70027 | 2/day | USA, AK | (70026, NWS, adjacent) |
| CAB | Cabauw | 06260 | 1/day | NL | AKA De Bilt |
| GRA | Graciosa | 08508 | 1/day | PT | Azores |
| LAU | Lauder | 93817 | 1/week | NZ | Later Vaisala software* |
| LIN | Lindenberg | 10393 | 4/day | DE | Lead centre |
| NYA | NyAlesund | 01004 | 1/day | NO | |
| PAY | Payerne | 06610 | 2/day | CH | Mainly Meteolabor* in 2017 |
| SGP | Southern Great Plains | 74646 | 4/day | USA, OK | AKA Lamont. |
| SOD | Sodankyla | 02836 | 2/day | FIN | |
| TAT | Tateno | 47646 | 2/day | JAP | Mainly Meisei* |
| TEN | Tenerife | 60018 | 2/day | ESP | |

*Table 1: GRUAN stations providing real-time reports on the GTS and NetCDF data in GRUAN archive (some of these are candidate GRUAN stations). * in the notes column indicates little or no public GRUAN data was available in 2017/18 for the reason given, but more data has since become available when new processing was certified. Some of the information came from J Tradowski (pers. comm., 2017).*

### 3.3. Differences between Vaisala RS92 reports at GRUAN and other stations

Is there a noticeable difference in quality between the Vaisala reports from GRUAN stations and other nearby stations using the same RS92 radiosonde? There might be extra care taken at GRUAN sites – although the highly automated Vaisala RS92 processing suggests that this would not be a major factor. There are extra ground checks at GRUAN sites, although these will not affect the Vaisala product from the ascents. The temperature ground check affects the GRUAN uncertainties but not the GRUAN temperatures themselves, the humidity ground check does affect the GRUAN humidities.

The German radiosonde network is relatively dense, good quality and in an area of relatively small background uncertainties so it is used to take a first look at station to station variations in accuracy. Because the Dutch GRUAN station Cabauw (CAB; WMO index 06260, it only reports at 00 UTC) is nearby it is also included. Figures 20 and 21 show individual station mean and root-mean-square (rms) O-B statistics for 2015/2016 data for 00 and 12 UTC respectively. For this period all the German stations were using RS92. The statistics cluster together quite tightly, the main exception being that one or two have an offset of about 8 m in the height (Z) statistics. The likely explanation is that the station height has been specified wrongly, certain stations elsewhere have biases of 20 m or more (Ingleby, 20017, section 3.4). In the upper troposphere/lower stratosphere there are two clusters for the relative humidity (RH) mean differences, with bigger differences at 12 UTC than 00 UTC. This is due to the different versions of the Vaisala processing (discussed in sections 2.4 and 3.1), the GRUAN stations are using the older processing which has larger biases compared to the background at 12 UTC in the stratosphere (but the background is not the truth). Six or seven of the stations do not report at 1000 hPa because of their elevation and it can be seen that at 925 hPa their temperature (T) rms differences are larger than for the other stations because of the proximity of the surface. At upper levels the temperature differences are slightly larger at 12 UTC than at 00 UTC, but this diurnal variation is clearer in the height differences. Mostly the results for Cabauw are very similar, although at 20 hPa the rms temperature differences are much larger, the stratospheric wind results for Cabauw also appear slightly worse (variation in background uncertainty might explain part of this).

It would be possible to try to extend this analysis more widely, to other countries. This raises the question of how much background and representation uncertainty vary geographically. Ingleby, 2017, looked at results for different latitude zones – there are features related to tropopause height. Ho et al. (2017) used reprocessed GPS-RO temperature retrievals for 2006-2014 to characterize radiosonde temperature biases and the variability of these biases in the upper troposphere and lower stratosphere for different radiosonde types. They compared Vaisala RS92 results for various different countries and found them to be generally consistent. There were some differences for Brazil, possibly because it is at lower latitude than the other countries compared, the sample size was also relatively small.

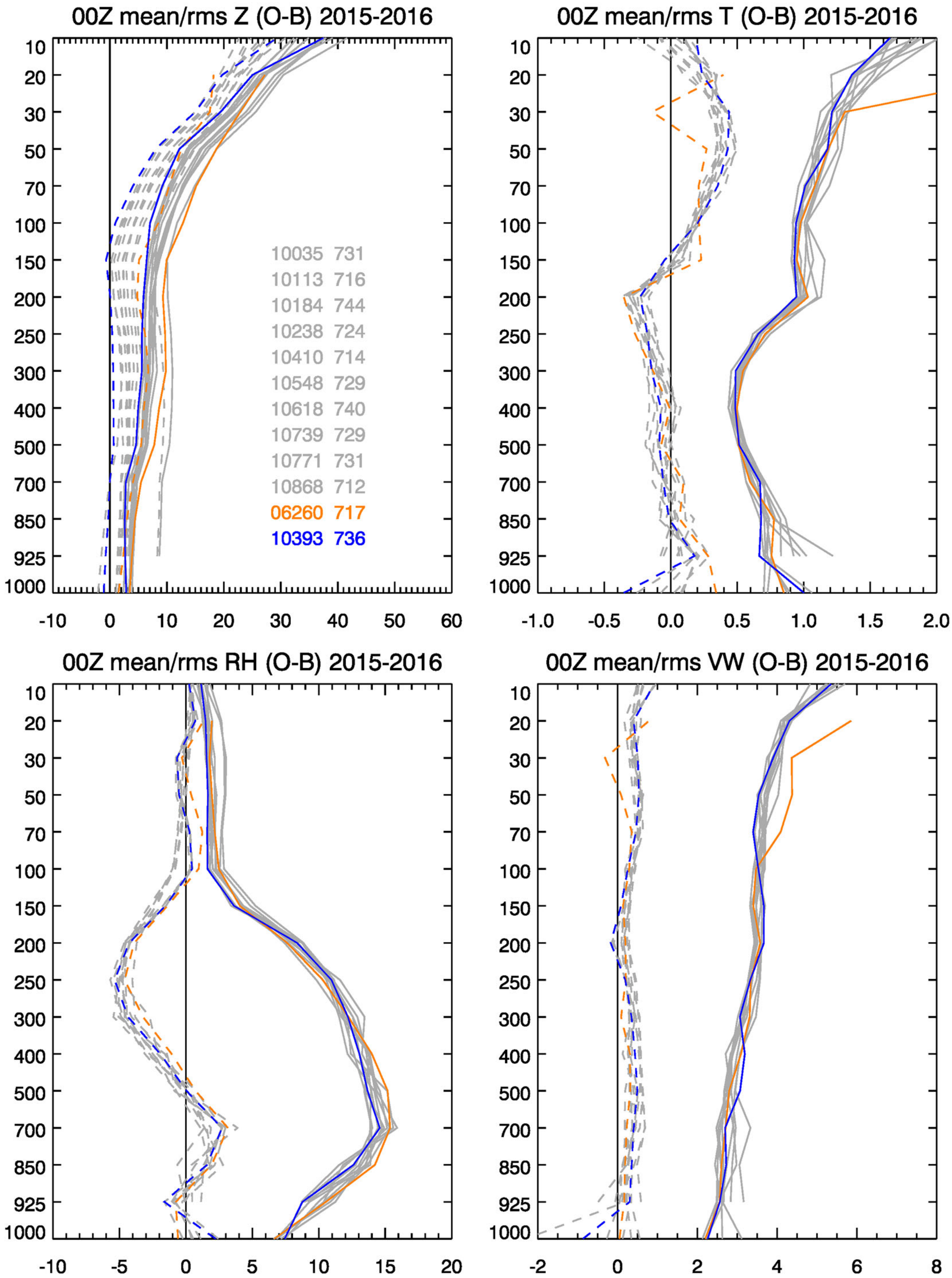



*Figure 20: Mean (dashed) and rms O-B statistics on standard levels for individual German and Dutch stations, 00 UTC TEMP reports 2015-2016, excluding values that failed the background check. Results for height (Z, m), temperature (T, K), relative humidity (RH, %) and wind (m/s: dashed mean speed difference, solid vector rms difference). WMO station identifiers and the number of reports (maximum at any standard level) are given in the key. Lindenberg – blue, Cabauw – orange, others – grey. The height and temperature statistics should be consistent (linked via the hydrostatic equation, with a minor contribution from humidity, but this has not been checked).*

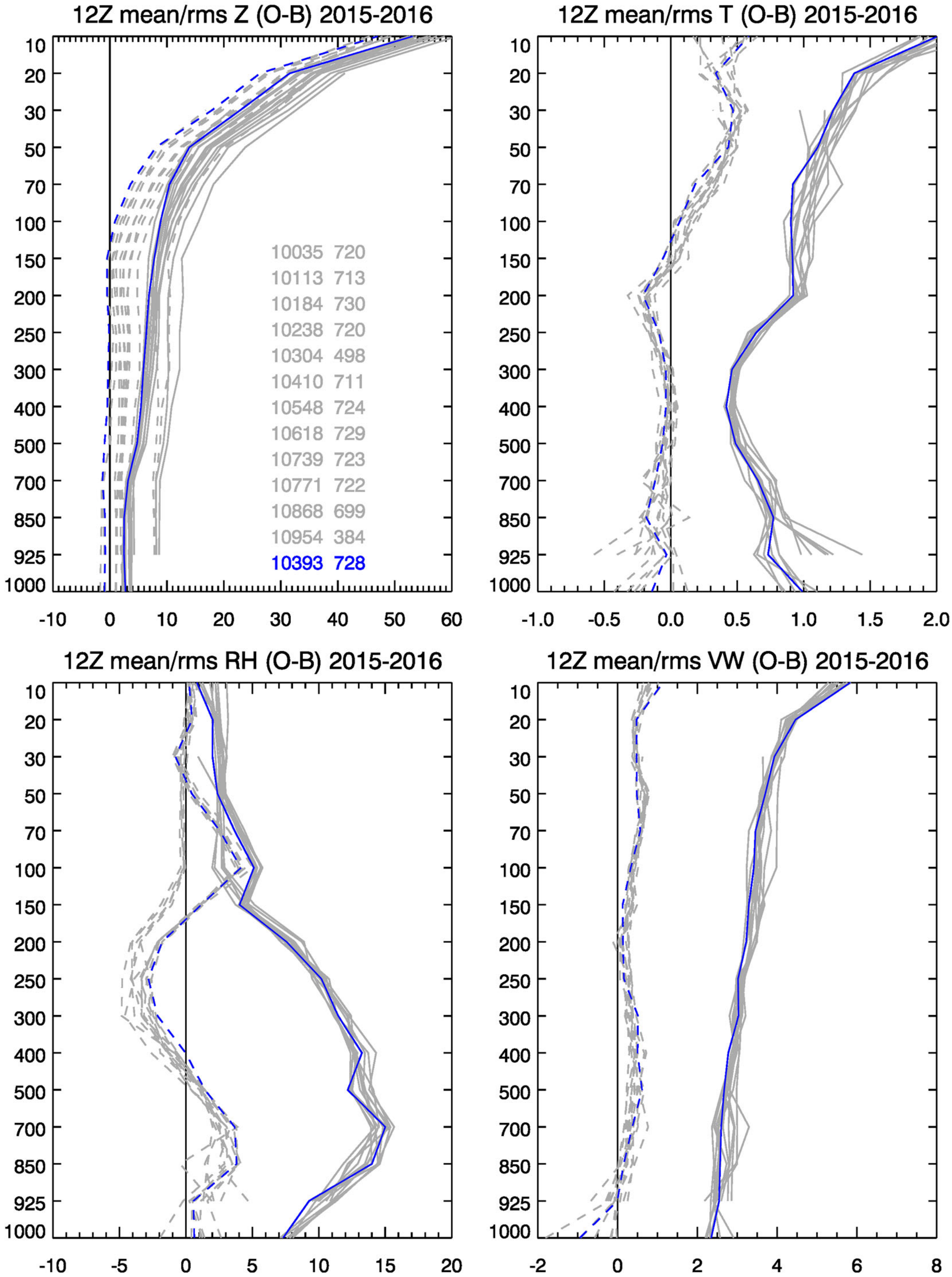



*Figure 21: As figure 20 but for 12 UTC reports (Cabauw does not usually report at 12 UTC).*

## 4. Discussion

Trying to estimate observation uncertainties and their correlations one inevitably finds links with 'representation (or representativeness) error', Lorenc (1986) identified this as the error in the observation operator. This has received a lot of interest recently including Frehlich (2011), Hodyss and Nicholls (2015) and Janjić et al. (2017). Much of the structure in O-B statistics, such as variations with latitude seen in Ingleby (2017), is due to background or representation uncertainty rather than measurement uncertainty.

High vertical resolution radiosonde data (often sampled every one or two seconds, about 5 or 10 metres) is the closest we have to completely resolved profiles and could be used to estimate vertical detail that cannot be represented on a particular model grid. Such profiles cannot help with horizontal aspects (for special purposes there can be repeat profiles ~1 hour later giving information on temporal variability and, under the Taylor hypothesis, spatial variability). Where a vertical profile is smooth and monotonic representation uncertainty is likely to be small. Larger representation uncertainties are likely near the surface (along with larger measurement uncertainties), around inversions (such as the boundary layer top and tropopause) and also where gravity waves are significant.

Gravity waves are partially represented in current forecast models but are often damped or artificially slowed (Shutts and Vosper, 2011; Preusse et al., 2014). They are visible in some stratospheric radiosonde profiles either temperature or wind (not usually both together) eg Ingleby (2017, Figures 4.3 and 4.17), occasionally they are well represented by the background but more often not. They are thus in a grey area. Kochin (2016) suggested filtering them out of radiosonde data but on balance it seems better to increase the representation uncertainty where they are present. This is an area for future work.

It is clearly easier to model uncorrelated uncertainties than correlated ones, and correlation does change the total information content of a vertical profile. However vertical uncertainty correlations can be modelled successfully. There are more subtle questions: are the uncertainties correlated with those from neighbouring radiosondes? Are they correlated with those from the same radiosonde one day, one month, six months or one year later? (The answer may depend on whether the radiosonde type or processing has changed in the meantime.) These latter questions are particularly important for climate studies.

Until about 2016 there was a proposal discussed at GRUAN meetings that a new BUFR template should be provided that would allow the reporting of GRUAN-style uncertainty estimates intended for use in data assimilation (DA). I was not in favour of this because I thought use of such estimates in DA was unlikely in the medium-term and because at remote stations, with a limit on message length, it might reduce the number of levels that were reported. It seemed likely that take up of the proposed template would be rather limited and more fundamentally the different priorities of GRUAN and DA cause problems. The role of representation error (section 1 and section 4 above) in DA means that the DA use of manufacturer/provider estimates of observation uncertainty has always been loose. Also, DA is less sensitive to changes in uncertainty than it is to small changes in temperature biases. The day/night variation in stratospheric temperature uncertainty is not currently included in DA systems, but appears to be only just above the noise level (Figures 20 and 21) and if required could be estimated and modelled within the DA system. In the ECMWF system temperature and humidity uncertainties are a function of radiosonde type – see Appendix. The GRUAN temperature uncertainty estimates have two aspects that appear undesirable for DA: the spikes and the apparently arbitrary difference between stations (both in Figure 18). One real aspect that is not captured by the GRUAN estimates is that there can be larger temperature and humidity errors just after emerging from cloud (the wet bulb effect, Dirksen et al., 2024) although this is less marked for Vaisala radiosondes than some others; the importance of modelling this for DA is not clear. GRUAN follows climate priorities in spending mush more time on temperature and humidity than on wind. In contrast, radiosonde winds have more impact within DA especially at high levels (Ingleby and Polichtchouk, 2025), in part this is due to the huge amount of temperature information available from satellites. GPS winds, as provided by most (but not all) radiosondes, are usually very good however there are questions about the high-frequency variability and it seems likely that the ‘pendulum motion’ filtering could be improved (Ingleby et al., 2022).

# 5. Summary

Estimates of observation uncertainty covariances have been computed for radiosonde standard level profiles using the method of Desroziers et al. (2005). Because complete assimilated profiles are needed for the chosen levels a large fraction of profiles may be discarded – hence most statistics are shown for 925 to 20 hPa and large subsets of radiosondes have been chosen. Particular attention has been paid to Vaisala RS92 and RS41 radiosondes, their observation uncertainties show small vertical correlations – very small for winds, at least for the standard levels sampled. For temperature there may be a small signal, particularly visible in the tropics, due to calibration uncertainty. The Vaisala upper tropospheric humidities have correlated uncertainties – reduced when recalculated just for the newer Vaisala RS41. The other features seem to be mainly due to other problems, such as a temperature bias in the extratropical background between about 100 and 20 hPa – measurement uncertainty due to solar radiation forms a small modification to this. One of the clearest features is a negative correlation between temperature and humidity at the same level – strongest around 850 and 700 hPa. This seems to be due to background uncertainty particularly in the vicinity of inversions (vertical displacement and/or smoothing – part might be classified as representation uncertainty). The magnitude and complexity of the background uncertainty near inversions is not well modelled by **B**. In the upper troposphere there is a minimum in temperature O-B differences, probably due to the relative lack of inversions there.

Other major radiosonde types have been examined more briefly. American LMS6 radiosondes perform well but have slightly higher off-diagonal elements particularly for humidity. They also appear to be more affected by 'competition' with aircraft data within the assimilation system. The vertical correlations calculated for Chinese wind profiles highlighted problems with wind direction offsets (radar orientation) at several Chinese stations. The results for Russian temperature profiles show problems which may be related to solar radiation effects and problems at low radar elevation angles. These diagnostics are quite sensitive and highlight various issues, but careful interpretation is needed to determine if the "issues" are related to the radiosonde data, the background fields or other aspects of the data assimilation system.

Comparing O-B statistics for Vaisala RS92 reports from individual German and Dutch stations, including two GRUAN stations the results are very similar, with slightly larger differences in the stratosphere. Particularly in the daytime some differences can be seen at upper levels between earlier and later versions of the humidity processing.

It seems reasonable to use average GRUAN v2 RS92 uncertainties as an estimate for operational RS92 and RS41 temperature and wind uncertainties, for upper tropospheric humidity operational RS92 uncertainties may be larger but the RS41 may well have smaller uncertainties. For temperature there is some station-to-station variation in the GRUAN uncertainties, the larger values should be taken. For temperature solar radiation and the difficulties of correcting for it mean that upper tropospheric and particularly stratospheric uncertainties are larger for daytime profiles. For data assimilation (and many other applications) these measurement uncertainties need to be supplemented by estimates of representation uncertainty.

All assimilation systems currently assume that radiosonde uncertainties are uncorrelated in the vertical – this study provides some support for this assumption, for most radiosondes at least. A subset of radiosonde profiles is subject to biases (in temperature, wind direction etc) or occasionally more complex forms of correlated errors. The approach taken is to try to detect and reject such correlated errors – this is partially successful, but could always be improved. The Russian vertical correlations of temperature (section 2.6) are moderate and there is no clear way to remove them. In the medium-term it would be useful to allow for vertically correlated errors as is done for some satellite soundings (e.g. Bormann et al., 2016). Perhaps more urgent is to specify observation uncertainties for major radiosonde types introduced since 2017 when the current uncertainties were set up (see Appendix).

## 6. Acknowledgements

Neils Bormann kindly provided his code for calculating and displaying the vertical covariances, some guidance in its use and comments on this manuscript. Lars Isaksen and Sean Healy provided support and comments. Michael Sommer provided the GRUAN data. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 640276 (GAIA-CLIM, Gap Analysis for Integrated Atmospheric ECV Climate Monitoring).

## 7. Appendix: radiosonde type dependence of uncertainty

In the ECMWF DA system, the same observation uncertainty used to be specified for all radiosondes, but profiles by radiosonde were introduced based on slightly smoothed versions of the results like those in Ingleby (2017). In practice we represent $\sigma_o$ as a function of pressure and radiosonde type for temperature and relative humidity, RH, wind uncertainties are taken as the same for all radiosondes. There is also some interaction with the "Huber norm" (non-Gaussian distribution) introduced by Tavolato and Isaksen (2015). In the first implementation (43r3, July 2017) of radiosonde type $\sigma_o$ there was a mistake in the Huber norm scaling factors which meant that effective temperature $\sigma_o$ values were somewhat larger than intended. Figure A1 shows the revised (45r1, June 2018) $\sigma_o$ values for temperature and RH.

Of the commonly used radiosonde types Ingleby (2017) found Vaisala RS92 and RS41, LMS6, Modem, Meisei and Shanghai to have the best performance for temperature and this is reflected, with slight variations, in figure A1. There is also a 'default' category (with similar $\sigma_o$ to the old profiles) for other radiosondes. The RH estimates were also updated and in this case all the new estimates are rather lower than the old ones (which were a function of temperature rather than pressure). The $\sigma_o$ values are specified on standard levels as shown and interpolated vertically to other reported levels. A practical issue is that as radiosonde types are replaced the list in the code should be updated (or new types get the default $\sigma_o$ which may be larger than we want).

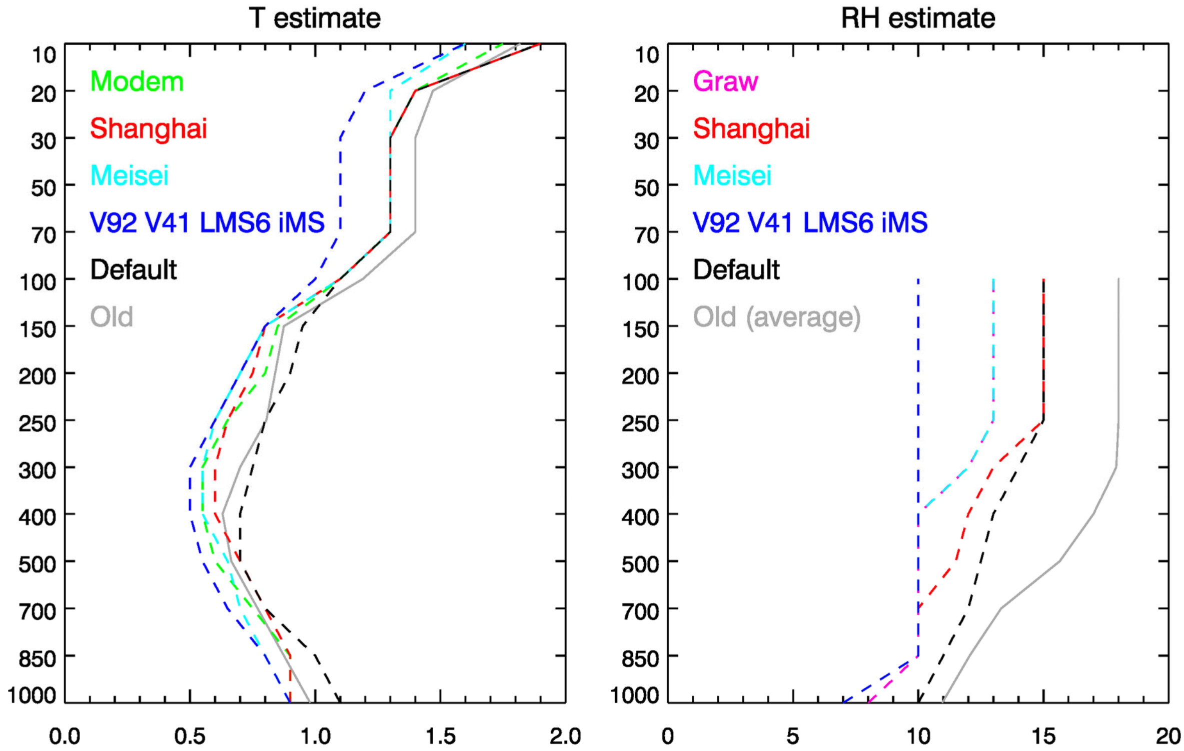



*Figure A1: Estimated observation uncertainties, $\sigma_o$, for temperature (K) and relative humidity (%RH) as a function of pressure (hPa) and radiosonde type (wind estimates are unchanged). NB. Large temperature increments are assimilated with reduced weight (through use of a Huber norm) rather than rejected. Very large departures are rejected. The temperature $\sigma_o$ values implemented in 2017 were larger than intended especially at low levels, this shows the corrected values. Before the 2017 change humidity $\sigma_o$ was a function of temperature. At levels above either 300 hPa or the -40°C isotherm only Vaisala humidities are assimilated, and no humidities are assimilated in the stratosphere. ("iMS" refers to the Meisei iMS-100, "Meisei" to other Meisei radiosondes.)*

## 8. References


Agustí-Panareda, A., Vasiljevic, D. et al., 2009: Radiosonde humidity bias correction over the West African region for the special AMMA reanalysis at ECMWF. *Quart. J. Roy. Meteor. Soc.*, 135, 595–617.

Bormann N, Bauer P. 2010. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. *Q. J. R. Meteorol. Soc*. 136: 1036-1050. https://doi.org/10.1002/qj.616

Bormann N, Bonavita M, Dragani R, Eresmaa R, Matricardi M, McNally A. 2016. Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Q. J. R. Meteorol. Soc.* 142: 1767-1780. https://doi.org/10.1002/qj.2774

Carminati F, W. Bell, S. Migliorini, S. Newman, A. Smith, 2016: An introduction to the GRUAN processor. Available from http://www.gaia-clim.eu/biblio/introduction-gruan-processor

Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background, and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* 131: 3385-3396.

Dirksen, R. J., M. Sommer, F. J. Immler, D. F. Hurst, R. Kivi, and H. Vömel, 2014: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmos. Meas. Tech.,* 7, 4463-4490, doi:10.5194/amt-7-4463-2014

Dirksen, R. J., et al. 2024: Report of WMO's 2022 Upper Air Instrument Intercomparison Campaign, 400 pp, WMO IOM Report No. 143 https://community.wmo.int/en/activity-areas/imop/publications-and-iom-reports

Frehlich, R. 2011: The definition of 'truth' for Numerical Weather Prediction error statistics. *Q.J.R. Meteorol. Soc.,* 137: 84-98. doi:10.1002/qj.738

Ho, S.-P., Peng, L., and Vömel, H. 2017: Characterization of the long-term radiosonde temperature biases in the upper troposphere and lower stratosphere using COSMIC and Metop-A/GRAS data from 2006 to 2014, *Atmos. Chem. Phys*., 17, 4493-4511, https://doi.org/10.5194/acp-17-4493-2017

Hollingsworth, A., Shaw, D.B., Lönnberg, P., Illari, L., Arpe, K., Simmons, A.J. 1986: Monitoring of observation and analysis quality by a data assimilation system. *Monthly Weather Review* 114, 861-879

Hollingsworth A, Lönnberg P. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I. The wind field. *Tellus* 38A: 111-136.

Hollingsworth, A., Lönnberg, P. 1989: The verification of objective analyses: Diagnostics of analysis system performance. *Meteorology and Atmospheric Physics*. 40, 3-27 doi.org/10.1007/BF01027466

Ide K, Courtier P, Ghil M, Lorenc AC. 1997. Unified notation for data assimilation. Operational, sequential and variational. *J. Meteorol. Soc. Jpn* 75: 181-189.

Ingleby, B., 2017: An assessment of different radiosonde types 2015/2016. *ECMWF Tech. Memo. 807*, 71 pp

Ingleby, B., 2022: Status and impact of radiosonde and surface observing systems, WMO TECO 2022 (in Topic 6) https://community.wmo.int/en/activity-areas/imop/publications-and-iom-reports

Ingleby, B., Rodwell, M. and Isaksen, L. 2016: Global radiosonde network under pressure. *ECMWF Newsletter* 149. Pp 25-30

Ingleby B, L. Isaksen, T. Kral, T. Haiden, M. Dahoui, 2018: Improved use of *in situ* data. *ECMWF Newsletter* 155

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, R. Weedon, 2016: Progress towards high-resolution, real-time radiosonde reports. *Bull. Amer. Meteor. Soc.*, **97,** 2149–2161

Ingleby, B., Motl, M., Marlton, G., Edwards, D., Sommer, M., von Rohden, C., Vömel, H., and Jauhiainen, H., 2022: On the quality of RS41 radiosonde descent data, *Atmos. Meas. Tech.* https://amt.copernicus.org/articles/15/165/2022/

Ingleby B. and Polichtchouk I, 2025: Stratospheric and tropospheric seasonality and its implications for observation requirements (in preparation)

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A. and Weston, P. (2017), On the representation error in data assimilation. *Q.J.R. Meteorol. Soc.*. doi:10.1002/qj.3130

Jensen, M. P., Holdridge, D. J., Survo, P., Lehtinen, R., Baxter, S., Toto, T., and Johnson, K. L.: Comparison of Vaisala radiosondes RS41 and RS92 at the ARM Southern Great Plains site, *Atmos. Meas. Tech.*, 9, 3115-3129, doi:10.5194/amt-9-3115-2016, 2016.

JCGM. Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. 2008. Available online: https://www.bipm.org/en/publications/guides/gum.html

Kats A., A. Balagourov, and V. Grinchenko, 2005: The impact of new RF95 radiosonde, introduction on upper-air data quality in the North-West region of Russia, TECO conference, available from http://ftp.wmo.int/pages/prog/www/IMOP/publications/IOM-82-TECO_2005/Posters/P2(07)_Russia_Kats.pdf

Kochin A. V., 2016: Correction of variation due to non-hydrostatic effects the observed temperature in upper-air sounding. WMO TECO conference, Madrid https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-125_TECO_2016/Session_4/P4(21)_Kochin_Correction.pdf

Krüger, K., Schäfler, A., Weissmann, M., and Craig, G. C.: Influence of radiosonde observations on the sharpness and altitude of the midlatitude tropopause in the ECMWF IFS, *Weather Clim. Dynam.*, 5, 491–509, https://doi.org/10.5194/wcd-5-491-2024, 2024.

Laroche S, & R. Sarrazin, 2013: Impact of Radiosonde Balloon Drift on Numerical Weather Prediction and Verification. *Weather and Forecasting*, **28,** 772-782

Lorenc, A.C. 1986 Analysis methods for numerical weather prediction. *Quart. J. R. Met. Soc*., 112, 1177-1194

Lorenc, A.C. 2007 : A study of o-b monitoring statistics from radiosondes, composited for low-level cloud layers. Forecasting Research Technical Report, NO. 504. Met. Office, Exeter (https://www.metoffice.gov.uk/learning/library search for FRTR_504_2007PT.pdf in online catalogue)

Miloshevich L.M., H. Vömel, D.N. Whiteman, and T. Leblanc. 2009: Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements. *J. Geophys. Res*., 114, D11305, doi:10.1029/2008JD011565.

Nash J., T. Oakley, H. Vömel, and Wei LI. 2011: WMO Intercomparison of High Quality Radiosonde Systems Yangjiang, China, 12 July - 3 August 2010, WMO Instruments and Observing Methods Report No. 107, available from http://www.wmo.int/pages/prog/www/IMOP/publications-IOM-series.html

Pauley P, Ingleby B (2022) Assimilation of in-situ observations. In: Park SK, Xu L (eds) *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV).* Springer. pp 293-371 in https://link.springer.com/book/10.1007/978-3-030-77722-7

Polichtchouk I., P. Bechtold, M. Bonavita, R Forbes, S.Healy, R. Hogan et al. 2021: Stratospheric modelling and assimilation. ECMWF Technical memorandum 877, DOI : 10.21957/25hegfoq

Preusse, P., Ern, M., Bechtold, P., Eckermann, S. D., Kalisch, S., Trinh, Q. T., and Riese, M., 2014: Characteristics of gravity waves resolved by ECMWF, *Atmos. Chem. Phys*., 14, 10483-10508, doi:10.5194/acp-14-10483-2014

Shutts, G. J. and Vosper, S. B., 2011: Stratospheric gravity waves revealed in NWP model forecasts, *Q. J. Roy. Meteor. Soc*., 137, 303-317, doi:10.1002/qj.7

Tavolato C. and L. Isaksen, 2015: On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. *Q. J. Roy. Meteor. Soc*., 141, 1514-1527, DOI: 10.1002/qj.2440

Waller JA, Dance SL, Nichols NK. 2016b. Theoretical insight into diagnosing observation-error correlations using observation-minus-background and observation-minus-analysis statistics. *Q. J. R. Meteorol. Soc.* 142: 418-431. https://doi.org/10.1002/qj.2661

Yu H., P.E. Ciesielski, J. Wang, H-C. Kuo, H. Vömel, and R. Dirksen, 2015: Evaluation of Humidity Correction Methods for Vaisala RS92 Tropical Sounding Data. *J. Atmos. Oceanic Technol*., 32, 397-411. doi: http://dx.doi.org/10.1175/JTECH-D-14-00166.1

WMO, 2023: Guide to Instruments and Methods of Observation (WMO-No. 8), https://community.wmo.int/en/guide-instruments-and-methods-observation-wmo-no-8