CONTRACT REPORT

Radiosonde assimilation experiments for Vaisala

June 2018

Author: Bruce Ingleby

Contract report for Vaisala Oyj

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

ECMWF

Series: ECMWF - Contract Report

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Abstract

Several assimilation-forecast experiments were performed in which aspects of radiosonde processing or usage were changed: 1. Making observation uncertainty a function of radiosonde type, 2. Removing the assimilation of upper-tropospheric radiosonde humidity, 3. Adding/subtracting an artificial 0.5K bias to radiosonde temperatures.

Introduction

In recent years several series of experiments were performed in which aspects of radiosonde processing or usage were changed. Some changes became operational, some did not perform well enough for operational implementation and others looked at the sensitivity to different aspects just to understand the system better. This report focuses on experiments performed as part of contract work, funded by Vaisala, which were: a) specifying radiosonde uncertainty to vary by observation type, b) denying upper tropospheric humidity data from radiosondes and c) looking at the effect of radiosonde systematic temperature biases.

One of the features of the observing system at this point of time is that radiosonde reports were/are migrating from alphanumeric TEMP format to binary BUFR format (Ingleby et al, 2016). At ECMWF some native BUFR reports, primarily from Europe, USA and Australia, are assimilated in place of TEMP reports. These mostly have high vertical resolution, and for assimilation are thinned to about 350 levels. They also report the time and position at each level – this facilitates the treatment of radiosonde drift, one of the changes discussed in this report. Unfortunately many BUFR reports have been reformatted from TEMP (still as separate parts), these are considered unusable. There is a gradual increase in the proportion of good, high-resolution BUFR reports; during the second half of 2017 most stations in the USA began sending high-resolution BUFR. A detailed investigations of radiosonde observation-minus-background statistics, split by radiosonde type was provided by Ingleby (2017).

Radiosonde bias correction

The operational ECMWF bias correction system for radiosonde temperature and humidity is described by Agustí-Panareda et al (2009). It uses the last 12 months of O-B statistics (updated once a month) to calculate bias corrections as a function of radiosonde type, pressure and solar elevation angle. The method assumes that any biases in the background fields are relatively constant. In general the temperature bias corrections are quite small, especially in the troposphere. Upper troposphere Russian humidities (not assimilated) have rather large bias corrections applied (Ingleby, 2017). Originally night-time RS92 were used as the reference, in June 2018, with IFS Cycle 45r1, ECMWF will change to using the average of night-time RS92 and RS41 data. The humidity correction for RS41 will also be switched off.

Some NWP centres and the ERA-Interim reanalysis do not bias correct radiosonde humidity. Some manufacturers have introduced "radiation corrections" for humidity. Newer versions of the RS92 radiosonde processing apply corrections to the humidity whereas older versions (still in widespread use in 2017) do not, making bias correction by radiosonde type problematic (see e.g. Yu et al, 2015). Humidity from the newer RS41 radiosonde is probably better than that from the RS92 data (Jensen et al, 2016), but the ECMWF automated system applied "bias corrections" of up to 4% RH to RS41 data in the upper troposphere, this is more a reflection of out-dated assumptions in the ECMWF bias correction scheme than problems with the radiosonde data. For these reasons it was decided to try completely switching off the radiosonde humidity bias correction in the ECMWF system (noRHBC experiment). Some other experiments were run at the same time also with no humidity bias correction. In retrospect this was slightly unfortunate because noRHBC verified worse against analyses than its Baseline experiment. However the incremental results relative to noRHBC should still give a good indication of whether a change is beneficial or not.

Notes on forecast verification

Forecast verification is a complex subject, partly because we never know the “truth” exactly. We have estimates of the truth from observations and analyses but each have their strengths and weaknesses – summarised in table 1. With its emphasis on the medium range ECMWF concentrates on verification against analyses. Because the ECMWF system is already very good most improvements only have a small effect on the verification scores, as discussed by Geer (2016). The usual scores examined are root-mean-square (rms) and anomaly correlation (ac); in some cases biases can make a significant contribution to the rms whereas they have less effect on the anomaly correlation.

	Versus Analyses	Versus Radiosondes
Coverage	Global ☺	Mainly continents. Smallish sample
Independence	Short range forecast not independent of analyses (this favours small analysis increments). In the tropics the problem can extend beyond D+2.	Generally considered independent ☺ (This assumption is more problematic for fields with high persistence e.g. SST.)
Biases	Can be a problem, less so for wind.	Some biases. Less of a problem for winds and tropospheric temperatures.
Sensitivity	Sensitive to small changes	Less sensitive, partly due to large representation uncertainty.

Table 1. Summary of the strengths and weaknesses of verification against analyses and against radiosondes.

Results

Specification of radiosonde uncertainties

Ingleby (2017) calculated O-B statistics for different radiosonde types and latitude bands for radiosonde standard levels. These were used to estimate revised estimates of observation uncertainty, σ_o, for temperature and relative humidity (RH), the wind estimates were left unchanged. The contribution of background uncertainty σ_b has to be removed:

$$rms(O-B)^2 \equiv \sigma_o^2 + \sigma_b^2 = \sigma_m^2 + \sigma_r^2 + \sigma_b^2$$

In practice σ_o has a contribution from representation uncertainty (σ_r) which may vary with latitude and synoptic situation as well as the measurement error (σ_m). According to Dirksen et al (2014) the measurement error has no direct dependence on latitude, but it does vary with pressure and, particularly for stratospheric temperature, with solar zenith angle. Difficulties in adjusting for solar radiation mean that high level temperatures have larger uncertainty during the day than at night. The split into observation, representation and background uncertainty is discussed further in Ingleby (2018, in preparation). In practice we choose to represent σ_o as a function of pressure and radiosonde type only for now (ignoring any dependence on solar elevation angle for now), and some vertical smoothing is applied to initial estimates. To a first approximation the representation uncertainty will be the same for different radiosonde types somewhat blurring differences due to measurement uncertainty. More elaborate models would need further work to generate statistics, coding to use them and further testing. There is also some interaction with the "Huber norm" (non-Gaussian distribution) introduced by Tavolato and Isaksen (2015). In the first implementation (43r3) of radiosonde type σ_o there was a mistake in the Huber norm scaling factors which meant that effective temperature σ_o values were somewhat larger than intended. Figure 1 shows the revised (45r1) σ_o values for temperature and RH. We specify these uncertainties even in cases where the values are not assimilated (such as for many upper tropospheric humidities – see next subsection).

Of the commonly used radiosonde types Ingleby (2017) found Vaisala RS92 and RS41, LMS, Modem, Meisei and Shanghai to have the best performance for temperature and this is reflected, with slight variations, in figure 1. There is also a 'default' category (with similar σ_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 σ_o values are specified on standard levels as shown and interpolated vertically to other reported levels. Within the ECMWF assimilation this issue applies to alphanumeric TEMP data, but not to high-resolution BUFR data – for the high resolution reports the levels assimilated are effectively chosen at random and significant levels have no special status.

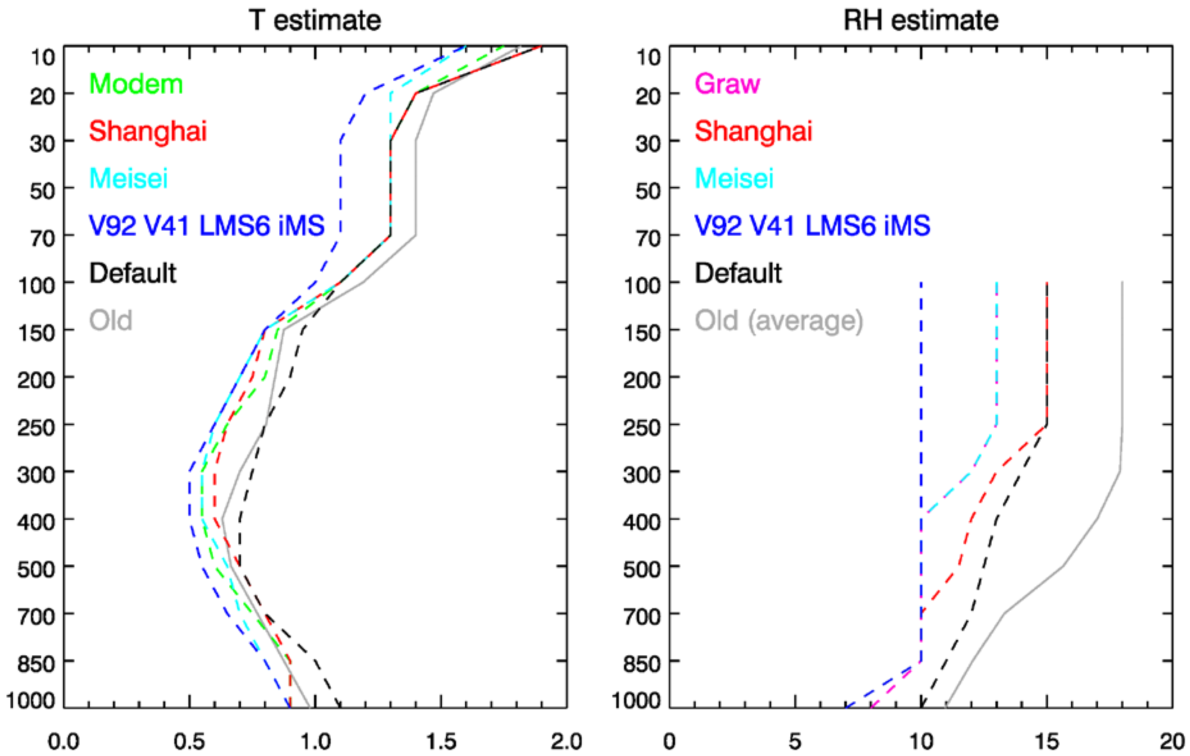


Figure 1. Estimated observation uncertainties, σ_o, for temperature (K) and relative humidity (%RH) as a function of pressure (hPa) and radiosonde type (wind estimates are unchanged). NB. Large temperature increments are assimilated with reduced weight (through use of a Huber norm) rather than rejected. Very large departures are rejected. The temperature σ_o values implemented in 2017 were larger than intended especially at low levels, this shows the corrected values. Before the 2017 change humidity σ_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.)

Figures 2 and 3 show verification scorecards for the initial implementation and then the corrected version (because of the practical constraints of testing these are for different period and different baseline systems). Figure 2 shows a generally positive impact at most forecast ranges (as discussed by Geer (2016) the different scores cannot be considered independent) and the change became operational in July 2017. The "correction" gave a more neutral impact (Figure 3) but it was decided to proceed with this change which should become operational in June 2018. Because they were tested relative to different model cycles the two changes were tested on different periods and cannot be compared directly.

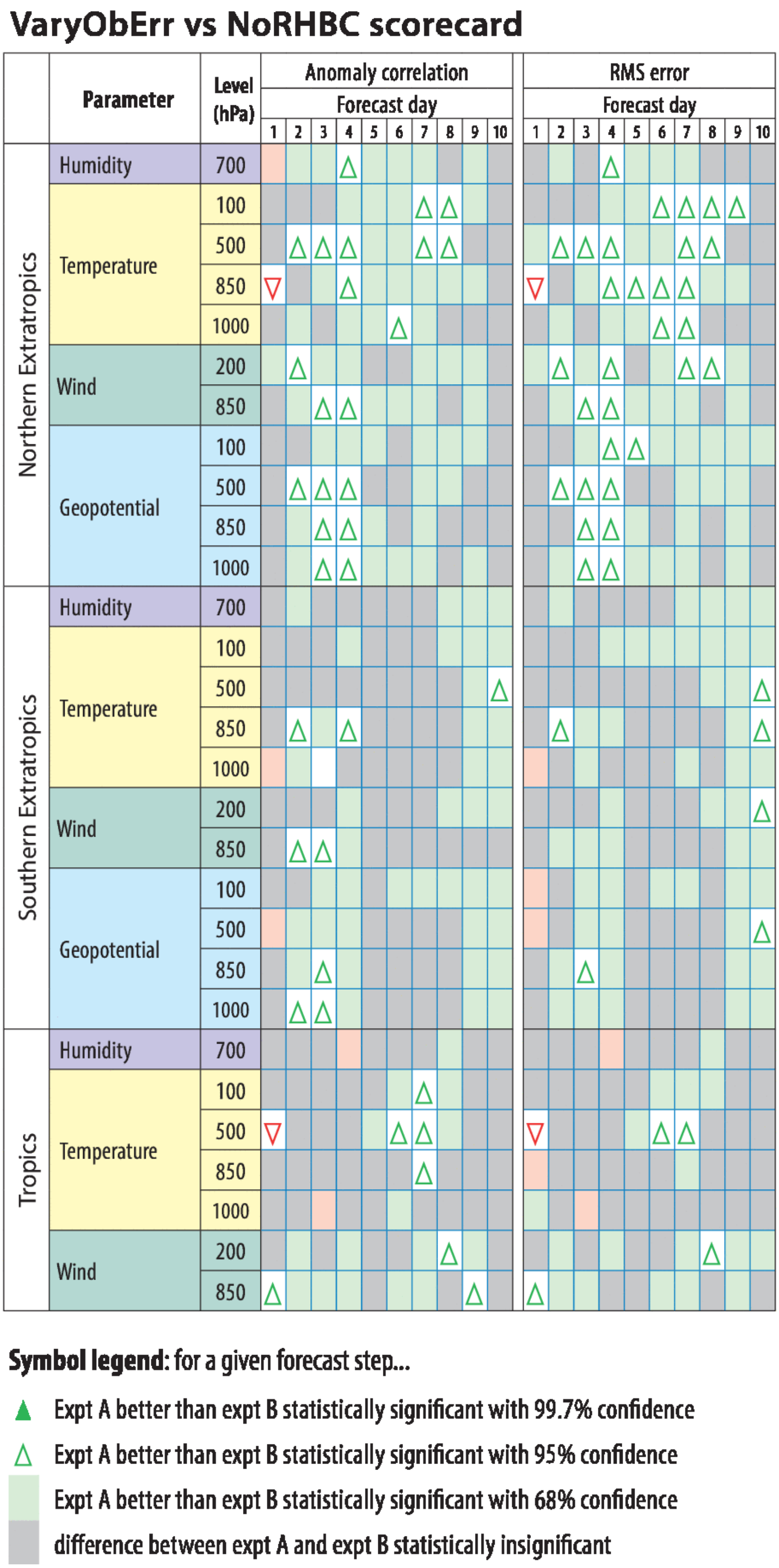


Figure 2. Scorecard for experiment varying σ_o as described in the text. Verification vs operational analyses for 2015-07-01 to 2015-09-30. The legend explains the green symbols, red symbols are similar but for experiment B better than A.

As explained in section 1.1 both experiments were run without bias corrections to radiosonde humidity but this "NoRHBC" change was not made operational.

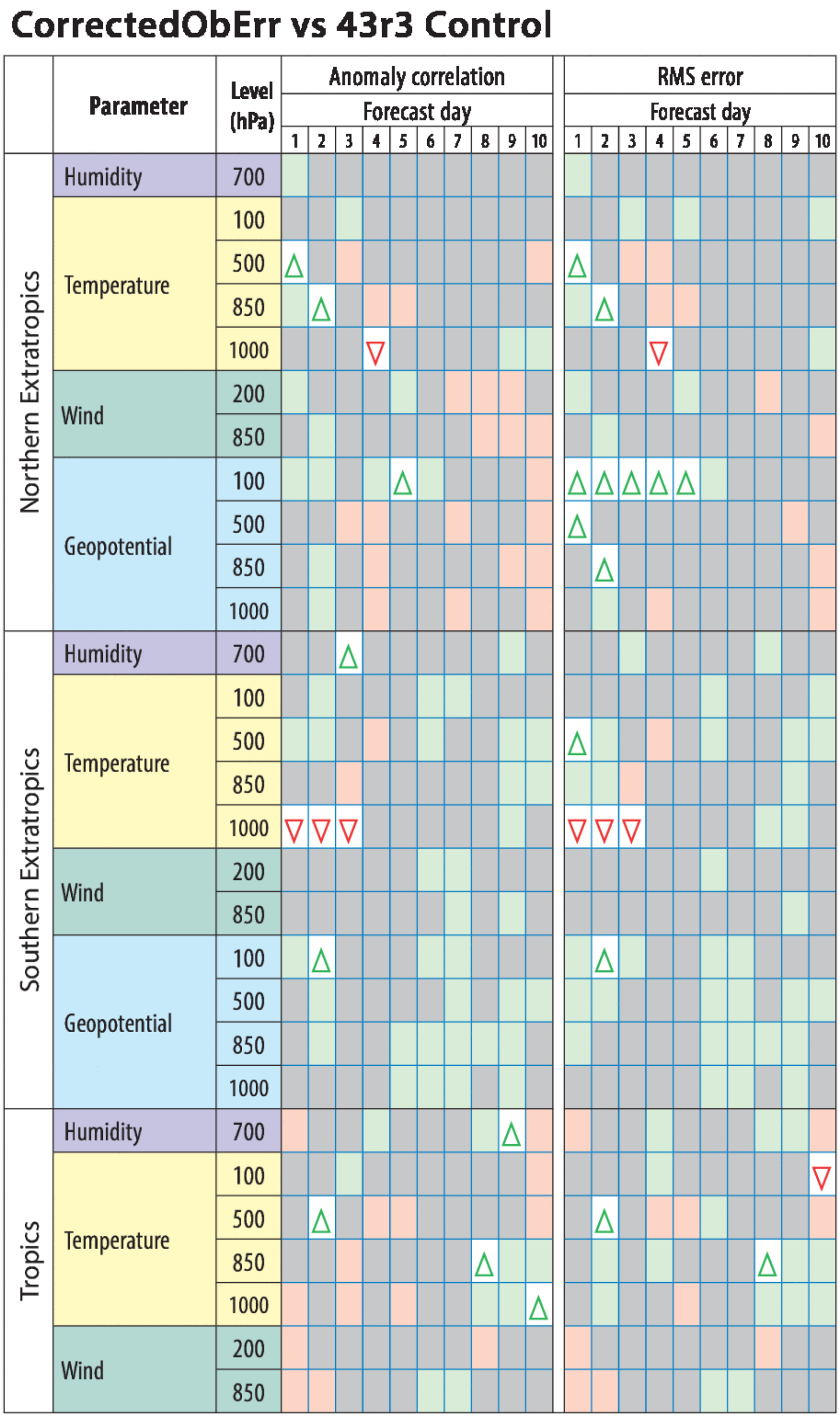


Figure 3. As figure 2, but for the correction to radiosonde σ_o. Trial period 2016-01-01 to 2016-03-31.

Impact of upper tropospheric humidity

Upper tropospheric humidity is particularly difficult to measure because a) capacitive humidity sensors react more slowly in cold conditions, b) some sensors are susceptible to contamination (icing) after passing through cloud. Solar radiation can also cause a problem during daytime. In general Vaisala RS92 and RS41 perform best (Nash et al, 2011; Ingleby, 2017) and their humidities are used up to 100 hPa or -80°C, whereas other radiosonde humidities are only used up to 300 hPa or -40°C. An experiment was run excluding upper tropospheric humidities, i.e. Vaisala radiosondes used the same limits as others. The scorecard is shown in figure 4. This suggests that the change is not statistically significant, but there are slightly more red (worse) scores. There is also a small degradation in 200 hPa radiosonde temperature rms(O-B) (not shown).

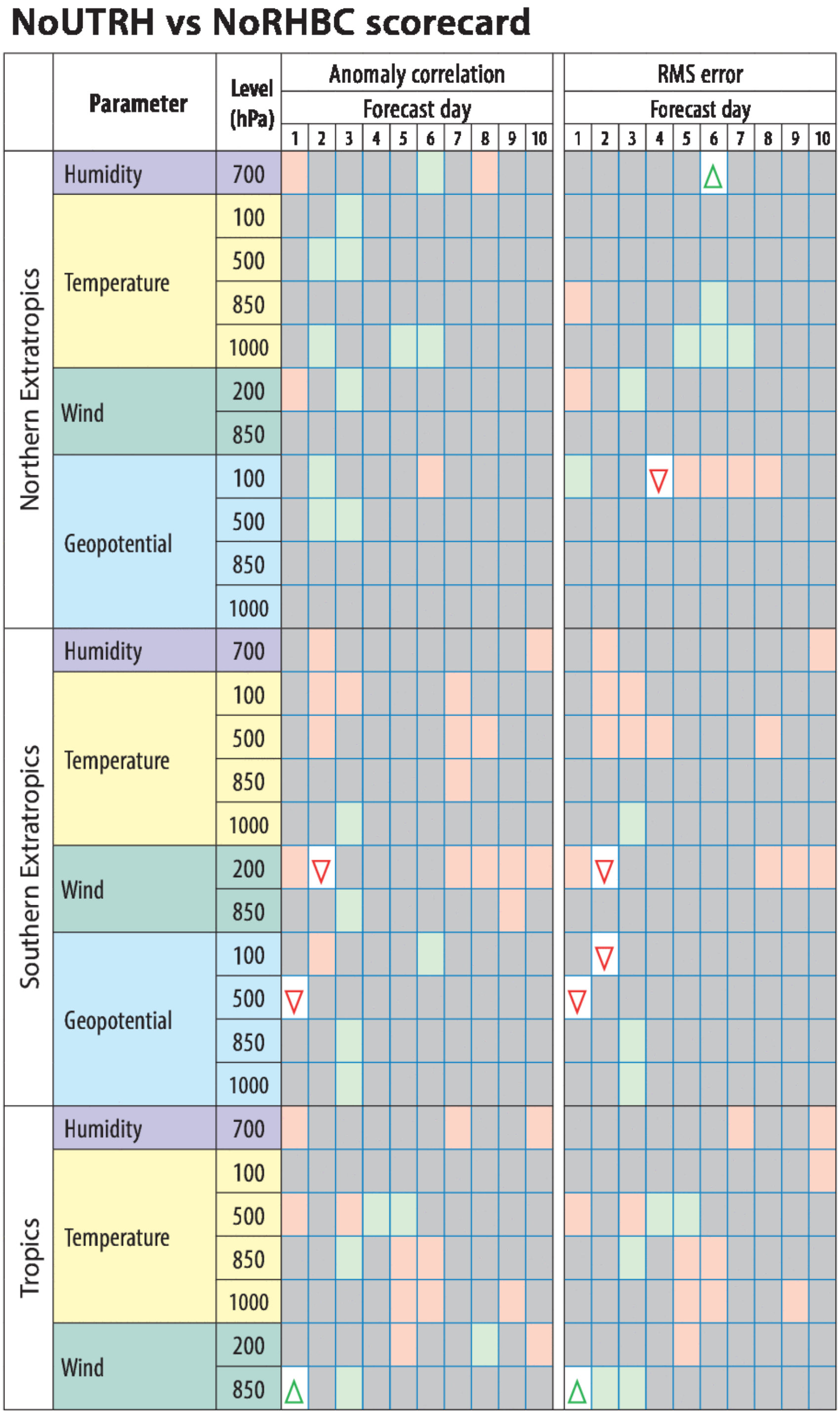


Figure 4. As figure 2 (same period and control) but excluding upper tropospheric radiosonde humidity.

Effects of radiosonde bias

Preliminary discussions considered adding random noise to radiosonde temperatures to look at the effect on the forecast system. However if the noise were completely uncorrelated then it seems likely that after about 12 hours forecast the impact would be very small. A problem with real radiosonde error profiles is that they may contain some vertically coherent structure – such as coming from solar radiation effects. (Recent work to try to estimate the vertical correlations of radiosonde uncertainty is ongoing.) Rather than choose a fairly arbitrary vertical scale it was decided to add a consistent offset (bias) to the radiosonde temperatures and a magnitude of 0.5 K was chosen – large enough to give a significant impact, but of a similar magnitude to radiosonde uncertainties (including representation uncertainty, see section 2.1; in the troposphere actual radiosonde temperature bias is unlikely to be larger than 0.1 or 0.2K). There was some concern that the impact might be asymmetric so two trials were run, one with 0.5K added to all radiosonde temperatures and the other with 0.5K subtracted from all temperatures: "Sonde+0.5" and "Sonde-0.5", respectively.

Figure 5 shows height anomaly correlation results for the two experiments relative to the control. The Sonde+0.5 experiment (red line) appears worse than the control, significantly so for some scores; its rms scores (not shown) are improved for 100 hPa height in the extratropics and 1000/850 hPa temperature in the tropics (probably due to changes in mean values), but most other rms scores are worse. In contrast the Sonde-0.5 experiment (black line in Figure 5) appears somewhat better than the Control for over half of the scores; in the northern extratropics the day two and three anomaly correlations look significantly improved, although from day seven onwards there are some degradations. The rms scores (not shown) are roughly opposite to those from Sonde+0.5 reinforcing the impression that they are dominated by differences in the mean fields. However it is still a surprise to see Sonde-0.5 apparently improving many verification scores, but this is mainly due to a cold bias in the forecast model.

Figure 6 shows the evolution of mean temperature fields over the ten day forecasts. The change in the analyses is less than the 0.5K applied to the observations but is more than 0.1K for some levels in the northern extratropics. In general the model is cooling in the troposphere over the forecast period (except for low levels in the northern extratropics) but the Sonde-0.5 experiment shows a bit less spin-down and hence appears better in verification against analyses. Geographical plots, as in Figure 7, show that the largest change in mean temperatures is around 500 hPa in northern mid/high-latitudes, especially over Europe. This is due partly to the relatively dense radiosonde network over Europe, but also to the provision and use of high-resolution radiosondes there (high-resolution radiosonde data was also assimilated over Australia during the trial period, but the horizontal density is lower). At low latitudes there is very little change to the mean temperatures – partly due to the sparseness of radiosondes in the tropics, but also probably due to more atmospheric physical factors (which tend to smooth out horizontal temperature gradients in the tropics). The relative geographical pattern suits some satellite channels (AMSU-A and others): their global standard deviation of O-B decreases slightly, although their bias corrections increase slightly (Geer, pers comm).

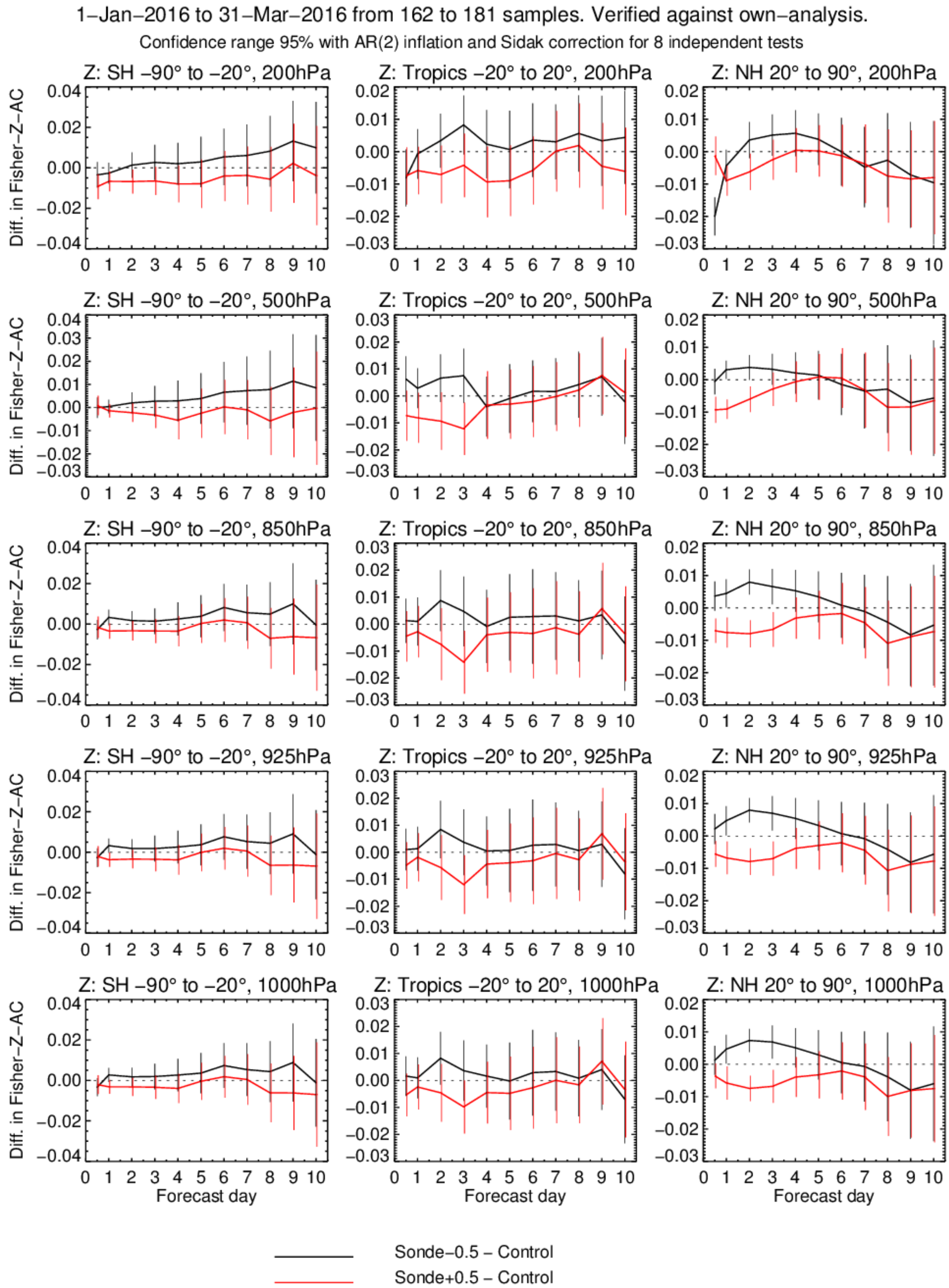


Figure 5. Height anomaly correlation differences for various tropospheric levels and three latitude bands, trial 2016-01-01 to 2016-03-31. Values above (below) zero suggest that the experiment is better (worse) than the Control – but see text for caveats.

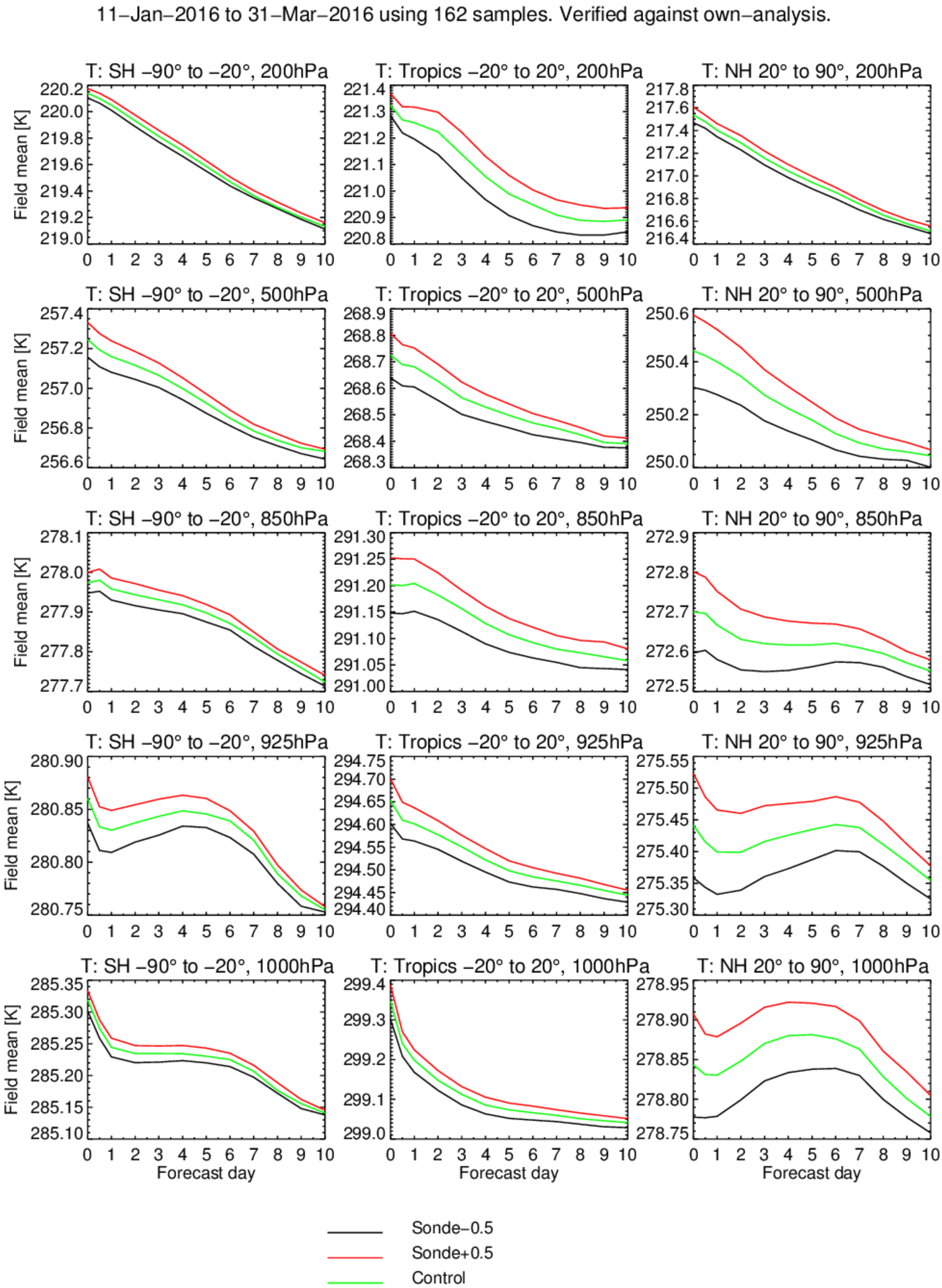


Figure 6. Evolution of mean temperature fields over the ten day forecasts.

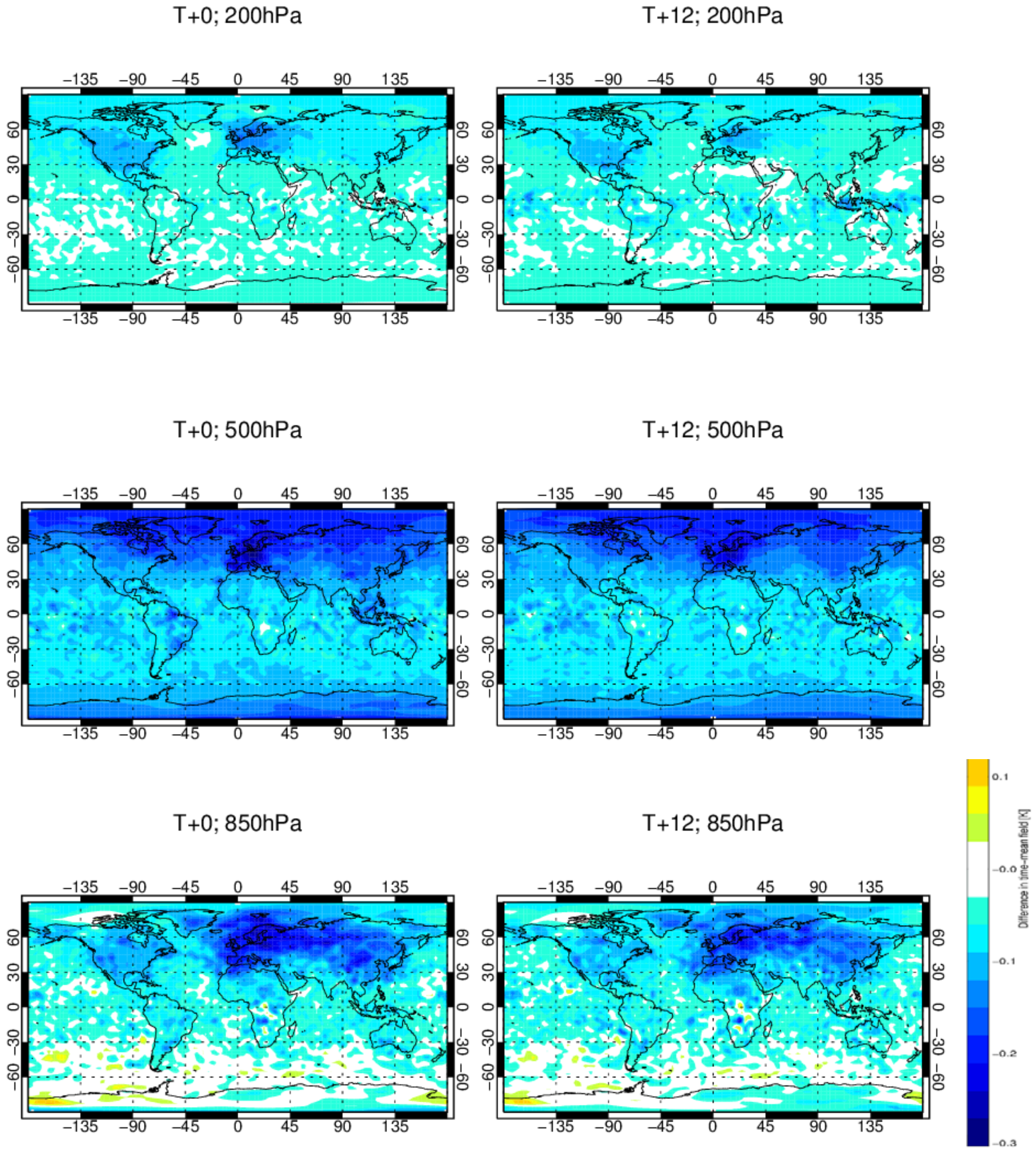


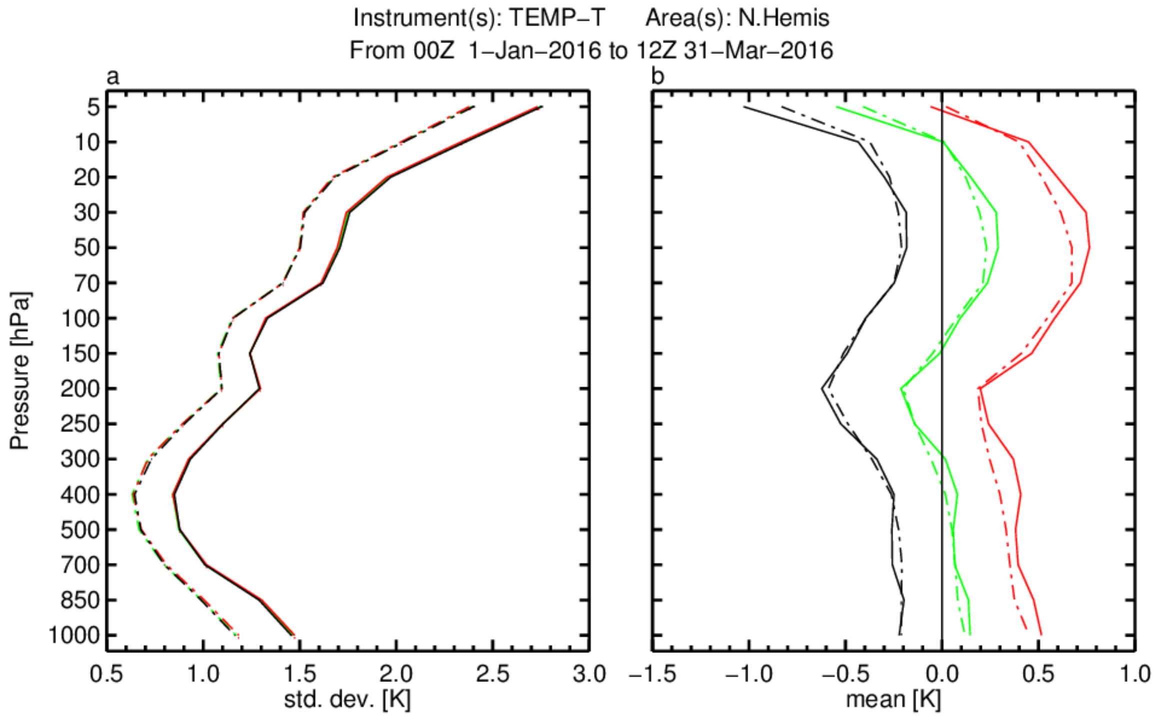
Figure 7. Difference in T+0 and T+12 temperature at three levels, Sonde-0.5 – Control.

We are used to verification against analyses providing misleading signals in the short-range (first two days) and beyond that in the tropics (hence Geer, 2016, omitted these from his study). It is somewhat alarming that for the Sonde-0.5 experiment verification against analyses is giving misleading signals into the medium range in the northern extratropics. This emphasizes the need for care in the interpretation of results and the need for “anchor” observations.

Eyre (2016) provides a useful discussion of bias correction issues in NWP: essentially one would like as many “anchor” observations as possible in order to constrain variational bias correction (VarBC) and similar schemes. In the ECMWF system satellite soundings (see Han and Bormann, 2016, and references) and aircraft temperatures are bias corrected using VarBC: leaving GPS-RO and radiosondes as the only observations anchoring atmospheric temperatures (other centres also bias correct these observations although the details may vary). Although bias correction is applied to radiosonde temperatures as described in section 1.1 there is an independent reference (not just analysis or background fields) so they can be considered as anchor observations – although it would be good if some radiosonde types could be assimilated without any bias correction applied. This has happened inadvertently to a limited extent – due to a technical issue BUFR reports with radiosonde type 99 or higher do not currently have a bias correction applied. Mostly these are modern types with small biases such as RS41, but the technical issue will be corrected so that we can choose which types to correct or not.

Starting with the noRHBC experiment described in section 1.1 there have been several attempts to switch off bias corrections for some radiosonde subsets – sometimes combined with rejections of extra Russian humidity data (eg above 400 hPa). These attempts have been ‘shelved’ because of minor degradations in verification against analysis (along with O-B statistics this is the main factor in implementation decisions). More use of verification against observations may be needed to better assess such changes to bias correction, both against in situ observations and others (Dahoui et al, 2017). Some tools are available at ECMWF, but they are not in general use for research experiments.

In general changes that increase analysis increments make verification against analysis look worse in the short range. An example can be seen in Figure 3 for 1000 hPa temperature in the southern extratropics – reduced temperature σ_o at low levels, means larger analysis increments there – at longer range there are some signs of improvement.



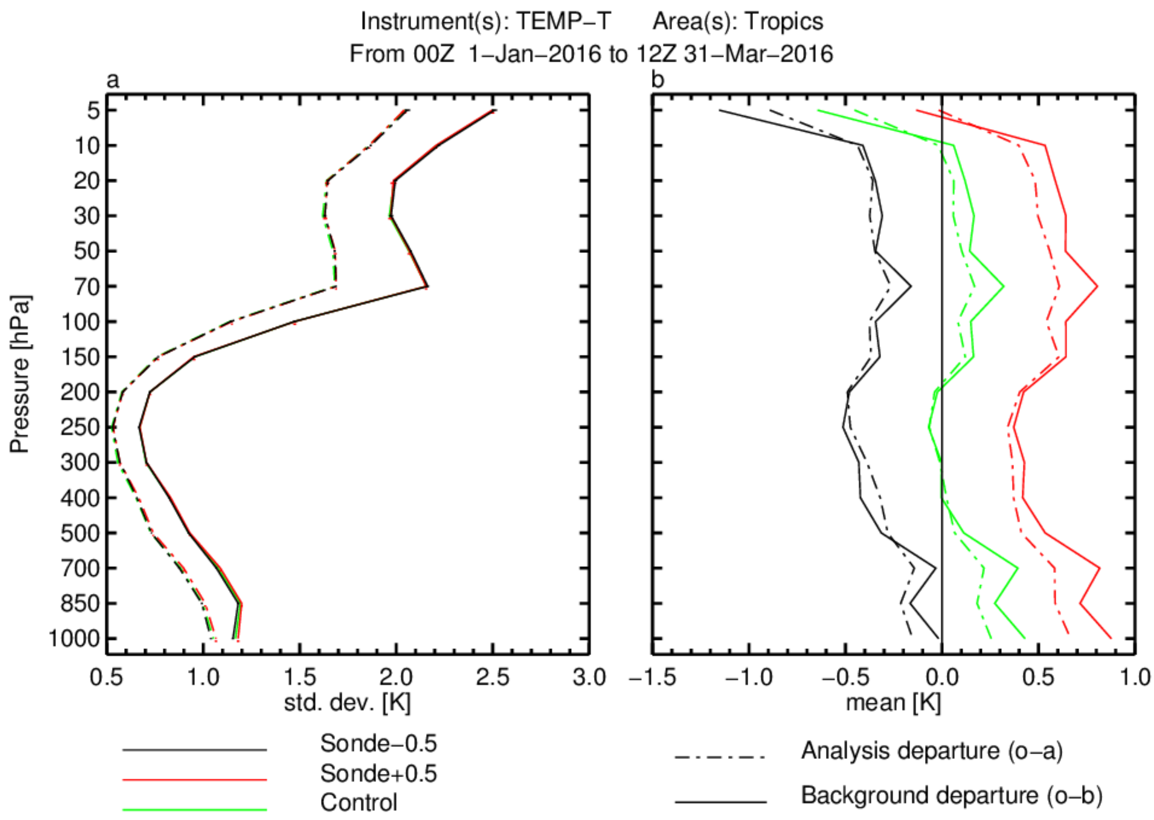


Figure 8. O-B radiosonde statistics for northern extratropics (top) and tropics (bottom). Note that the mean background values are rather similar, it is the O values that change between experiments.

Summary

Three issues were investigated.

1. Whether specification of observation uncertainty, σ_o, by radiosonde type is beneficial? The revision of σ_o based on the results of Ingleby (2017) did produce forecast improvements and was made operational in July 2017 (with a more minor correction due to be made in June 2018), this included a general reduction in humidity σ_o as well as making temperature and humidity σ_o a function of radiosonde type.
2. How much benefit comes from the use of upper tropospheric humidity from radiosondes? The impact on forecasts of removing the upper tropospheric humidities was quite small. Upper tropospheric humidity is however important for estimation of climate radiative feedbacks.
3. What effect does degradation of radiosonde temperatures – simulated by adding/subtracting a 0.5K bias – have? In terms of the impacts on forecast scores these were quite large changes, partially because the same change was applied to all radiosondes. But the results were surprising in that subtracting 0.5K improved various scores against analysis; adding 0.5K gave a clear degradation.

The modelling of observation uncertainty is still relatively simple (but extra complexity has to be justified). There is a case for separating the estimates of measurement uncertainty and representation uncertainty, but for consistency this would ideally be done for all observation types, it then becomes a major change.

The misleading signals from verification against analyses of the Sonde-0.5 experiment are a cause of concern as far as the assessment of forecast performance goes. In this case the slight cooling fits model biases and the spatial pattern seems to suit some satellite channels. This case shows the continuing need for anchor observations like radiosondes, and for the care needed in observation bias correction. It also shows the need for verification against observations.

Acknowledgements

Lars Isaksen supervised this work and contributed useful discussions. Tomas Kral made technical changes to the observation preprocessing to allow two of the changes. Both the IVER package (maintained by Alan Geer) and Verify were used to look at forecast performance. The publication quality versions of the scorecards were provided by Anabel Bowen. We would like to thank Aki Lilja, Vaisala Oyj for useful discussions and Vaisala Oyj for funding this study.

References

Agustí-Panareda, A., D. Vasiljevic and Coauthors, 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.

Dahoui, M, L. Isaksen, G. Radnoti, 2017: Assessing the impact of observations using observation-minus-forecast residuals. ECMWF Newsletter No. 152. 27-31

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

Geer A.J. 2016: Significance of changes in medium-range forecast scores, Tellus A: Dynamic Meteorology and Oceanography, 68:1, DOI: 10.3402/tellusa.v68.30229

Eyre, J. R. (2016), Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties. Q.J.R. Meteorol. Soc., 142: 2284-2291. doi:10.1002/qj.2819

Han W. and N. Bormann, 2016: Constrained adaptive bias correction for satellite radiance assimilation in the ECMWF 4D-Var system. ECMWF Tech Memo 783.

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, 2016a: Progress towards high-resolution, real-time radiosonde reports. Bull. Amer. Meteor. Soc., 97, 2149–2161, doi: 10.1175/BAMS-D-15-00169.1.

Ingleby B. 2017: An assessment of different radiosonde types 2015/2016. ECMWF Technical Memorandum No. 807

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.

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

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

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