We would like to thank all three anonymous referees for their comments and constructive feedback. The comments and our responses are addressed below, where the referee's original comments are copied in italics. All page and line numbers in the referee's comments refer to the originally submitted version of the paper. All page and line numbers in our response refer to the revised paper.

Referee 1 (review received and published on 21 Feb. 2018)

Page 2, lines 1-3: it is not directly clear from the text what can be obtained from what. Could the authors reformulate this sentence?

Thank you for addressing this confusion. We have reworded the sentence on page 2:

"Refractivity is obtained by inverting bending angle profiles using the Abel transform. Refractivity is a function of temperature and water vapor pressure, therefore with auxiliary information (observations or model) of one, the other can be retrieved."

Page 6 equation (3): I suggest to use Greek phi instead of LAT for latitude and Greek lambda instead of LON for longitude, but this is a matter of taste.

Thank you for the suggestion. We decided to keep LAT and LON instead of phi and lambda since this notation is more explicit and clear to those not familiar with the Greek symbol terminology. In addition, LAT and LON only appear in this one equation so the notation does not become cumbersome.

Referee 2 (review response received and published on 1 Mar. 2018)

The authors tested two methods on reducing spatial sampling errors in the comparison of profiles of collocated radiosonde and GPSRO. They concluded that the "double differencing" method (by using a NWP model data) will most effectively reduce the sampling errors. And if that method is not available, restricting the comparisons to within an ellipse parallel to the wind flow will also reduce the sampling errors caused by atmospheric variability. I like this paper. But the authors discussed only the impact on root-mean-square (rms) errors, not mentioning at all the impact of the mean differences between GPSRO and radiosonde. This authors should add plots or tables to indicate how different methods impact the mean difference…that is very important to studies of utilizing RO data to analyze radiosonde accuracy.

Thank you for the suggestion. We have computed the mean and standard deviation differences of the three variables for all stations, and these figures are now included in Appendix A. Our results indicate that the mean is not sensitive to co-location or sampling differences, as all six co-location methods result in similar mean profiles for all variables across all stations. This is most likely due to cancelation of differences of opposite signs during the computation of the mean. However, we see the impact of the different co-location criteria on the standard deviations, which are very similar to the RMS

differences. The mean differences also give some indication of the biases in the RO minus radiosonde differences, as discussed in the new Appendix A.

Referee 3 (review received and published on 7 Mar. 2018)

1. Abstract: Pls. provide the time period of the data used.

We have added "Using data from 2011-2014, …" to the abstract.

2. P2, L26-30: Those are assumptions. Are there any prior studies on this or can the authors conduct some sensitivity studies on this?

Daily operational weather analyses (e.g. those published by the National Centers for Environmental Prediction at http://mag.ncep.noaa.gov/model-guidance-model-parameter.php?group=Model%20Guidance&model=GFS&area=NAMER&ps=model) show that, in general, the winds in the free atmosphere (above the boundary layer) tend to blow parallel to contours (isopleths) of temperature and water vapor, the former through the thermal wind relationship and the latter through stretching and shearing deformation. We state on page 3 that it is our hypothesis that the spatial variability of refractivity will be minimized along the direction of wind flow (i.e. within the ellipse parallel to wind flow). Kitchen (1989) hints at this concept in terms of temperature. In our paper, Fig. 2 on page 6 provides strong visual evidence for our hypothesis in the ERA-Interim refractivity fields. In addition, we conducted experiments using only ERA-Interim data, applying the ellipse method to ERA-interim refractivity and eliminating temporal differences by using ERA-interim grid-point data co-located at the same time. We mention at the beginning of Section 3 that applying the ellipse method using ERA-Interim data showed a significant reduction in RMS refractivity differences relative to both circles; however, we chose not to show these results for brevity.

3. Some reviews on prior studies considering different approaches to match pairs spatially, such as Mears et al. (2015)

Thank you for the reference. We have added the following paragraph on page 2 and included the reference there:

"Kitchen (1989) compared RS with infrared (IR) soundings and noted that sampling errors generally dominate the total error associated with the comparison. Bruce et al. (1977) compared RS with satellite retrieved temperature, also discussing the impacts of sampling errors. Mears et al. (2015) applied a bilinear best fit plane to microwave radiometer observations to reduce sampling errors during comparisons with ground-based GPS observations. Fasso et al. (2014) and Ignaccolo et al. (2015) both propose a statistical modeling approach to reduce sampling errors associated with RS inter-comparisons and balloon drift analysis. In this study, we focus on RO-RS comparisons, and attempt to reduce the sampling errors that occur during these comparisons."

We have also added references on the spatial-temporal correction method, which is also called the "double-differencing" method. These additions are on pages 3 and 8 of the revised paper.

4. Section 2.1: Pls. provide references for COSMIC and GRUAN projects. This applies to other places too. For GRUAN, pls. use Bodeker et al (2016) on
https://www.gruan.org/documentation/articles/

We have added a reference to the COSMIC and other RO missions on page 3. We have also added references to the GRUAN project on page 4 and in the caption of Fig. 1 on page 4.

5. P6, L25-30, add this reference for spatial correction applied to RS and space-borne MWR comparisons: Mears, C. A., J. Wang, D. Smith, and F. J. Wentz (2015), Intercomparison of total precipitable water measurements made by satellite-borne microwave radiometers and ground-based GPS instruments, J. Geophys. Res. Atmos., 120, doi:10.1002/2014JD022694.

We added this reference on page 2 where we think it is most appropriate.

6. Fig.2: How many co-located RO profiles do you get within big circle, small circle and ellipse given your time matching window?

The numbers of co-located RO profiles are given in panel (a) of Figs. 5, 6, and 7, as well as in Fig. 4.

If you get more than one, do you average them and then compare with RS?

No, as with other comparison studies of this type (Wickert et al., 2004; Kuo et al., 2005; He et al., 2009; Xu et al., 2009; Ho et al., 2010; Sun et al., 2010; Zhang et al., 2011; Wang et al., 2013; Vergados et al., 2014), we compute RO-RS differences for each individual RO-RS pair (Eq. (4)): for a single RS at a given time and pressure level, there are (often) multiple co-located RO profiles, and each individual pair (pairing the co-located RO profile with the RS) is considered separately when the differences are computed. We have added the following explanation to page 7 of the revised draft:

"For each RS at a given time and pressure level, RO profiles are co-located to the RS under the time and geometric constraints discussed above. There can (and often are) multiple co-location pairs with the same RS at a given time and pressure level."

7. Section 3.1: Table 1 shows that the reduction in RMS due to filtering is the largest for Eclipse. Is it due to smaller spatial coverage of Eclipse or the fact that the RS & RO profiles matched in Eclipse are more representative of the same air mass?

Thank you for this interesting observation. In Table 1 on page 9, the largest overall percent reduction in RMS due to filtering is within the small circle, with very similar reductions in RMS within the ellipse. Since the sampling errors are reduced within the small circle and ellipse (relative to the large circle), vertical representativeness errors become more significant, therefore we see a larger percentage reduction in vertical

representativeness errors in the small circle and ellipse relative to the circle when filtering. We have added this analysis on pages 9-10:

"Within the large circle, filtering has less of an impact on reducing vertical representativeness percent errors since sampling errors tend to dominate (mostly due to large spatial differences). The RO-RS percent differences in the small circle and ellipse are more affected by vertical representativeness errors since the sampling errors are relatively small."

8. Main goal of applying filtering and different spatial matching is to reduce spatial and temporal sampling errors and increase the signals of instrument errors for both instruments. It would be very useful if the authors can provide one example showing this. For example, how the comparisons show one instrument's know errors more clearly after applying the new matching approaches proposed here.

The goal of this paper is only to develop and test methods of reducing representativeness and sampling errors during RO-RS co-location comparisons. Our methods do not isolate the errors of the individual observation systems, they only reduce the representativeness and sampling errors so the error analysis can be performed more accurately. To discuss the errors of the individual observation systems after applying this method is important but is outside the scope of this paper. Anthes and Rieckh (2017) discuss a method of isolating errors using the so-called 3-cornered hat method, which may address your question.

9. One of unique things about GRUAN radiosonde data is its uncertainty estimate. How does your algorithm reduce ï A¸s, then give you a better estimate of RO uncertainty if you already know the uncertainty of RS data in (see Immler et al. 2010 and other GRUAN references on www.gruan.org) for this.

We are aware of the error uncertainty estimates; they are a very nice feature provided by GRUAN. Our two methods of reducing sampling errors caused by horizontal inhomogeneities in the atmosphere do not address uncertainty in the RS or RO observations. This important question is outside the scope of this paper. However, we have added plots comparing the mean differences between the RO and RS in a new Appendix A, and this comparison does provide some information on the biases in the RO minus RS differences. This is discussed in the new appendix.

10. Read this: Fassò, A., Ignaccolo, R., Madonna, F., Demoz, B. B., and Franco-Villoria, M.: Statistical modelling of collocation uncertainty in atmospheric thermodynamic profiles, Atmos. Meas. Tech., 7, 1803-1816, doi:10.5194/amt-7-1803-2014, 2014.

Thank you for the reference, we have added this reference to a new paragraph on page 2.

References:

Anthes, R. and Rieckh, T.: Estimating observation and model error variances using multiple data sets, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2017-487, 2017.

He, W., Ho, S.-P., Chen, H., Zhou, X., Hunt, D., and Kuo, Y.-H.: Assessment of radiosonde temperature measurements in the upper tro- posphere and lower stratosphere using COSMIC radio occultation data, Geophys. Res. Lett., 36, https://doi.org/10.1029/2009GL038712, 2009.

Ho, S.-P., Zhou, X., Kuo, Y.-H., Hunt, D., and Wang, J.: Global evaluation of radiosonde water vapor systematic biases using GPS radio occultation from COSMIC and ECMWF analysis, Remote Sensing, 2, 1320–1330, https://doi.org/10.3390/rs2051320, 2010.

Kitchen, M.: Representativeness errors for radiosonde observations, Quart. J. Roy. Meteor. Soc., 115, 673–700, 1989

Kuo, Y.-H., Schreiner, W., Wang, J., Rossiter, D., and Zhang, Y.: Comparison of GPS radio occultation soundings with radiosondes, Geophys. Res. Lett., 32, https://doi.org/10.1029/2004GL021443, 2005.

Sun, B., Reale, A., Seidel, D. J., and Hunt, D. C.: Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics, J. Geophys. Res., 115, https://doi.org/10.1029/2010JD014457, 2010.

Vergados, P., Mannucci, A., and Ao, C.: Assessing the performance of GPS radio occultation measurements in retrieving tropospheric humidity in cloudiness: A comparison study with radiosondes, ERA-Interim, and AIRS data sets, J. Geophys. Res., 19, 7718–7731, https://doi.org/10.1002/2013JD021398, 2014.

Wang, B.-R., Liu, X., and Wang, J.-K.: Assessment of COSMIC radio occultation retrieval products using global radiosonde data, Atmos. Meas. Tech., 6, 1073–2013, https://doi.org/10.5194/amt-6-1073-2013, 2013.

Wickert, J., Schmidt, T., Beyerle, G., Konig, R., and Reigber, C.: The radio occultation experiment aboard CHAMP: Operational data analysis and validation of vertical atmospheric profiles, J. Meteor. Soc. Japan, 82, 381–395, https://doi.org/10.2151/jmsj.2004.381, 2004.

Xu, X., Luo, J., and Shi, C.: Comparison of COSMIC radio occultation refractivity profiles with radiosonde measurements, Adv. Atmos. Sci., 26, 1137–1145, https://doi.org/10.1007/s00376-009-8066-y, 2009.

Zhang, K., Fu, E., Silcock, D., Wang, Y., and Kuleshov, Y.: An investigation of atmospheric temperature profiles in the Australian region using collocated GPS radio occultation and radiosonde data, Atmos. Meas. Tech., 4, 2087–2092, https://doi.org/10.5194/amt-4-2087-2011, 2011.

Reducing representativeness and sampling errors in radio occultation–radiosonde comparisons

Shay Gilpin [1], Therese Rieckh [1,2], and Richard Anthes [1]

[1]COSMIC Program Office, University Corporation for Atmospheric Research, Boulder, CO
[2]Wegener Center for Climate and Global Change, University of Graz, Austria

Correspondence to: Shay Gilpin (sgilpin@ucar.edu)

Abstract. Radio occultation (RO) and radiosonde (RS) comparisons provide a means of analyzing errors associated with both observational systems. Since RO and RS observations are not taken at the exact same time or location, temporal and spatial sampling errors resulting from atmospheric variability can be significant and inhibit error analysis of the observational systems. In addition, the vertical resolutions of RO and RS profiles vary and vertical representativeness errors may also affect the comparison. ~~Typically in~~ In RO–RS comparisons, ~~RS~~ RO observations are co-located with ~~RO~~ RS profiles within a fixed time window and distance, i.e. within $3-6$ h and circles of radii ranging between $100-500$ km. In this study, we first show that vertical filtering of RO and RS profiles to a common vertical resolution reduces representativeness errors. We then test two methods of reducing horizontal sampling errors during RO–RS comparisons: restricting co-location pairs to within ellipses oriented along the direction of wind flow rather than circles, and applying a ~~spatial~~ spatial-temporal sampling correction based on model data. ~~We~~ Using data from $2011-2014$, we compare RO and RS differences at four GCOS Reference Upper-Air Network (GRUAN) RS stations in different climatic locations, in which co-location pairs were constrained to a large circle (~666 km radius), small circle (~300 km radius), and ellipse parallel to the wind direction (~666 km semi-major axis, ~133 km semi-minor axis). We also apply a ~~spatial~~ spatial-temporal sampling correction using European Centre for Medium-Range Weather Forecasts Interim Reanalysis (ERA-Interim) gridded data. Restricting co-locations to within the ellipse reduces root mean square (RMS) refractivity, temperature, and water vapor pressure differences relative to RMS differences within the large circle, and ~~either reduce or~~ produces differences that are comparable to or less than the RMS differences within circles of similar area. Applying the ~~spatial~~ sampling correction shows the most significant reduction in RMS differences, such that RMS differences are nearly identical with the ~~spatial~~ sampling correction regardless of the geometric constraints. We conclude that implementing the ~~spatial~~ spatial-temporal sampling correction using a reliable model will most effectively reduce sampling errors during RO–RS comparisons, however, if a reliable model is not available, restricting spatial comparisons to within an ellipse parallel to the wind flow will reduce sampling errors caused by horizontal atmospheric variability.

1 Introduction

Radio occultation (RO), a relatively new method of atmospheric measurements, has established itself as an important atmospheric observational system. By measuring the phase delay of radio-waves sent from Global Positioning Satellites (GPS)

traversing ~~horizontally~~ quasi-horizontally through Earth's atmosphere to low-Earth orbiting satellites, RO obtains accurate and precise vertical profiles of bending angles (Melbourne et al., 1994). Refractivity is obtained ~~from~~ by inverting bending angle profiles using the Abel ~~Transform, and from refractivity, stratospheric temperature and tropospheric temperature and water vapor can be retrieved (where tropospheric temperature and water vapor require auxiliary measurements or modeldata) . RO~~

5 ~~provides observations with global coverage, high vertical resolution, little or no instrument bias, and bending angles which can be directly~~ transform. Refractivity is a function of temperature and water vapor pressure, therefore with auxiliary information (observations or model) of one, the other can be retrieved. Either bending angles or refractivity may be assimilated into numerical weather prediction models ~~(Melbourne et al., 1994; Kursinski et al., 1997; Anthes et al., 2008).~~(Eyre, 1994).

Since the proof-of-concept GPS/MET mission in 1995 (Ware et al., 1996), RO profiles of refractivity, temperature, and water

10 vapor have been compared to radiosondes (RS) to assess the quality of RO retrievals and the performance of RS instruments. RS are considered a standard for comparison due their long history of in situ measurements~~starting in 1958 (Ho et al., 2017)~~, and several studies have used RS as a reference for RO retrieval analysis (Wickert et al., 2004; Kuo et al., 2005; He et al., 2009; Xu et al., 2009; Ho et al., 2010; Sun et al., 2010; Zhang et al., 2011; Wang et al., 2013; Vergados et al., 2014). Conversely, due to RO's properties of high accuracy and precision, high vertical resolution, and global coverage, RO has been used to

15 evaluate the performance of various RS instruments. For example, Kuo et al. (2005) demonstrated that RO observations are of high enough accuracy and resolution to differentiate between RS instruments and ~~asses~~ assess their performance, particularly instrument biases due to geographic region, radiation errors, day/night biases, etc.

One of the main difficulties associated with RO–RS comparisons comes from temporal and spatial differences between nearby RS and RO soundings. Since both measurements are not taken at the exact same time or location, temporal and spatial

20 errors (sampling errors) can be a significant part of the computed RO–RS differences. To reduce the effects of sampling errors, the majority of previous studies have restricted co-located RO observations to within a fixed time range and distance, typically within $3-6$ h of the RS launch and within circles of radii ranging from $100-500$ km centered at the RS launch site. Alternatively, Staten and Reichler (2009) used smaller radii circles ($3-36$ km) and fitted a second order polynomial to the root mean square (RMS) differences in order to filter out atmospheric variability. Weather-scale atmospheric variability within

25 these circles and time ranges is the major cause of these sampling errors (Sun et al., 2010)~~, and in this studywe~~.

Kitchen (1989) compared RS with infrared (IR) soundings and noted that sampling errors generally dominate the total error associated with the comparison. Bruce et al. (1977) compared RS with satellite retrieved temperature, also discussing the impacts of sampling errors. Mears et al. (2015) applied a bilinear best fit plane to microwave radiometer observations to reduce sampling errors during comparisons with ground-based GPS observations. Fasso et al. (2014) and Ignaccolo et al. (2015) both

30 propose a statistical modeling approach to reduce sampling errors associated with RS inter-comparisons and balloon drift analysis. In this study, we focus on RO–RS comparisons, and attempt to reduce ~~these errors~~ the sampling errors that occur during these comparisons.

We apply two methods to reduce sampling errors caused by atmospheric variability in RO–RS comparisons. First, we restrict co-location pairs to within ellipses oriented along the direction of wind flow rather than circles. ~~Orienting ellipses along the~~

35 ~~direction of wind flow should minimize atmospheric variability due to the interdependence of atmospheric composition and~~

~~wind.~~ Temperature and water vapor gradients in the free atmosphere tend to be perpendicular to wind flow, resulting in refractivity (a function of both temperature and water vapor pressure) gradients to also approximately lie perpendicular to wind flow. Therefore, we hypothesize that the spatial variability of refractivity within ellipses of semi-major axis a oriented along the wind direction will be reduced compared to the variability within circles of radius a. The second method consists of applying

5   a ~~spatial~~ spatial-temporal sampling correction to the RO–RS co-location pairs ~~based on~~ using the "double-difference" method (Chander et al., 2013; Tradowsky et al., 2017). In the double-difference method, each data set is compared to an intermediate reference data set; in our study we use the European Centre for Medium-Range Weather Forecasts Interim Reanalysis (ERA-Interim) model data ~~(see www.ecmwf.int/era).~~ (Dee et al., 2011, and https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-i By subtracting the ~~model data from both the~~ reference model data at the corresponding RO and RS locations in space and time

10   from the RO and RS observations, the spatial-temporal sampling errors are largely eliminated, leaving mainly the ~~background and representativeness errors are removed, leaving only the~~ RO and RS observational errors. ~~The approach has been called the "double-difference" method (Haimberger et al., 2012; Wong et al., 2015; Tradowsky et al., 2017).~~ In addition to the sampling error, the different vertical resolution of the RO and RS profiles can lend to representativeness errors when compared. To reduce these errors, we filter the profiles to a common vertical resolution before the comparisons.

15   Although this paper considers RO–RS comparisons, the vertical filtering and methods of reducing sampling errors can be applied to comparisons of other data pairs, such as any two sounding systems or observations and models. However, the amount of filtering and the geometric constraints on the co-location pairs may have to be adjusted for different comparisons. For example, comparisons of RO or RS soundings with infrared (IR) or microwave (MW) soundings, which have much different vertical resolutions, would require a greater filtering of the RO or RS profiles to make them comparable to the lower-resolution

20   profiles.

The first section describes the data sets, filtering methods, and methodology implemented in this study. Next, we discuss aspects of the ellipse co-location method conducted using the Lindenberg RS station. The following section describes the results of co-locations using both the ellipse and ~~spatial~~ sampling correction methods at four different RS stations. In the final section we summarize the results and discuss further impacts and implications, followed by an appendix which includes mean

25   and standard deviation differences and further discussion of the spatial-temporal sampling correction.

2   Methods

2.1   RO and RS data sets

All RO profiles are provided by the COSMIC Data Analysis and Archive Center (CDAAC), which can be found at http://cdaac-www.cosmic.ucar.edu. RO refractivity, temperature, and water vapor pressure profiles are taken from the wetPrf profiles pro-

30   vided on a uniform $100\,\mathrm{m}$ mean sea level height grid. All available processed, high-quality profiles from the COSMIC-1, GRACE 1 and 2, and Metop-A/B missions during the time periods of comparisons are used. Summaries of these missions can be found at the CDAAC website above. The temperature and water vapor pressure profiles are computed using a 1D-VAR

method for moisture retrievals. Details on the 1D-VAR wet temperature and water vapor retrievals can be found at http://cdaac-www.cosmic.ucar.edu/cdaac/doc/documents/1dvar.pdf.
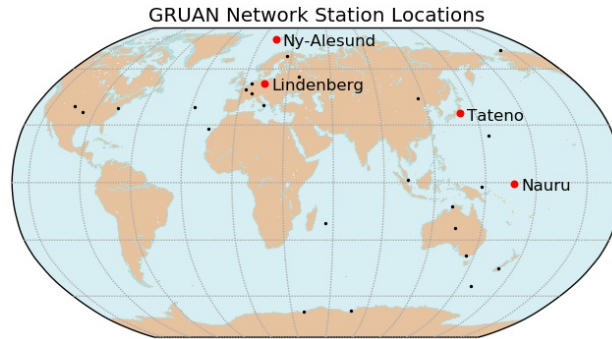
Figure 1. Map of the GRUAN RS network. Labeled locations with red dots are RS stations used in this study, black dots mark the other GRUAN network RS stations. For more information about these and other GRUAN stations, see https://www.gruan.org.

All RS profiles are provided by the Global Climate Observing System (GCOS) Reference Upper-Air Network ~~(GRUAN)~~ (GRUAN; see Bodeker et al., 2016, for more information on the GRUAN project) and downloaded from the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center. GRUAN RS were chosen for their reference-quality observations and well documented uncertainties (Seidel et al., 2009; Immler et al., 2010; Seidel et al., 2011), allowing for a better analysis of the reduction of sampling errors associated with each co-location method.

We chose four stations in different climates for this study: Lindenberg, Germany (LIN); Ny-Ålesund, Norway (NYA); Tateno, Japan (TAT); and Nauru, Nauru (NAU) for the following time periods: 2014, 2013, 2012, and ~~2011-2013, respectfully~~ $2011-2013$, respectively. (Nauru is the only station in which the full period of activation was used; this is due to the low number of RS launches during 2011 through late August 2013.) Figure 1 is a map of the GRUAN RS stations with the locations of the four stations used in this study labeled and marked in red. All four stations use the Vaisala RS92 radiosonde instruments (for details on GRUAN RS processing see Dirksen et al., 2014).

For refractivity comparisons, RS refractivity is computed under the assumption of a neutral atmosphere (Smith and Weintraub, 1953):

$$N = 77.6\frac{\cancel{P}\;p}{\cancel{T}\;T} + 3.73 \times 10^5 \frac{e}{T^2} \tag{1}$$

where N is refractivity ($N - $ units), $\cancel{P}\;p$ is dry air pressure (hPa), T is temperature (Kelvin) and e is water vapor pressure (hPa).

We generate two data sets: an unfiltered data set which contains all original RO and RS profiles, and a filtered data set containing the vertically filtered versions of the original RO and RS profiles.

2.2 Filtering

Representativeness errors result from two different aspects of the RO and RS observations: (1) GRUAN RS have a temporal resolution of $1\,\text{s}$, with vertical resolution of $5-10\,\text{m}$ on average (Ladstädter et al., 2015), which is much finer than the $100\,\text{m}$ vertical resolution of the RO wetPrf profiles, and (2) RS observations are a series of point measurements, whereas RO observations are weighted averages of a cylindrical volume of atmosphere with horizontal scales of $150-300\,\text{km}$ (Kursinski et al., 1997; Kuo et al., 2004; Anthes, 2011). Vertical filtering of both RO and RS profiles should decrease the representativeness errors caused by differences in ~~vertical resolutions and observation types~~observation type and vertical resolutions (Lohmann, 2007), resulting in a more meaningful comparison between profiles. Removal of structures with very small vertical scales also has the effect of reducing representativeness errors associated with different horizontal scales (footprints) of the observations (Kitchen, 1989). Structures with very small vertical scales in the RS profile are likely associated with horizontal scales too small to be resolved by the RO observations. Though the majority of previous RO–RS comparison studies do not filter the RO or RS profiles before comparison, Kuo et al. (2004) filtered the profiles to remove structures with vertical scales less than $1\,\text{km}$ before comparison with model analyses in an effort to minimize vertical ~~sampling differences~~representativeness errors.

To remove small-scale, unrepresentative structures in both the RO and RS profiles, we applied the Savitzky-Golay low-pass filter (Savitzky and Golay, 1964). We first linearly interpolated profile variables (refractivity, temperature, and water vapor pressure) to a $10\,\text{hPa}$ ($\sim100\,\text{m}$) uniform vertical grid, then filtered ~~using a~~the full profile using a fixed $40\,\text{hPa}$ ($\sim400\,\text{m}$) filter window and quadratic fitting polynomial. We tested various combinations of filter windows and number of passes to determine the effects of filtering on the RS and RO profiles (not shown). The number of passes of the same filter result in minor smoothing effects relative to different filter windows, and filter windows larger than $40\,\text{hPa}$ cause too much smoothing of the RS profile. RO profiles, due to their lower resolution relative to the RS profiles, ~~demonstrated~~showed little change under different filter windows~~, therefore~~. Therefore, $40\,\text{hPa}$ was chosen as a sufficient window to remove ~~small scale~~small-scale features in the RS profile while preserving the overall structures of both the RS and RO profiles. To accommodate for the differences in vertical resolution, RO profiles underwent a single pass of the $40\,\text{hPa}$ filter and RS profiles underwent three passes of the $40\,\text{hPa}$ filter.

2.3 RO–RS co-location criteria and statistics

We co-located RO and RS observations using the ellipse method every $10\,\text{hPa}$ between $1000-10\,\text{hPa}$ at each RS location. We included RO occultations taken within $3\,\text{h}$ of the RS launch at Ny-Ålesund, Tateno, and Nauru and within $1\,\text{h}$ at Lindenberg due to the high number of RS launches (at least four times daily). We considered three different geometric constraints to co-locate RO profiles: (1) large circle of radius $6\,^\circ$ latitude ($\sim666\,\text{km}$), (2) small circle of radius $2.6\,^\circ$ latitude ($\sim300\,\text{km}$), and (3) ellipse parallel to the wind direction, $6\,^\circ$ latitude semi-major axis and $1.2\,^\circ$ latitude ($\sim133\,\text{km}$) semi-minor axis, each centered at the RS location. (The ellipse and small circle are circumscribed by the large circle, and the small circle was chosen such that the area within the small circle and ellipse are approximately the same, see Fig. 2). The X and Y coordinates of the points on the circles and ellipses are constructed using the following parameterization (a: semi-major axis, b: semi-minor axis, for circles
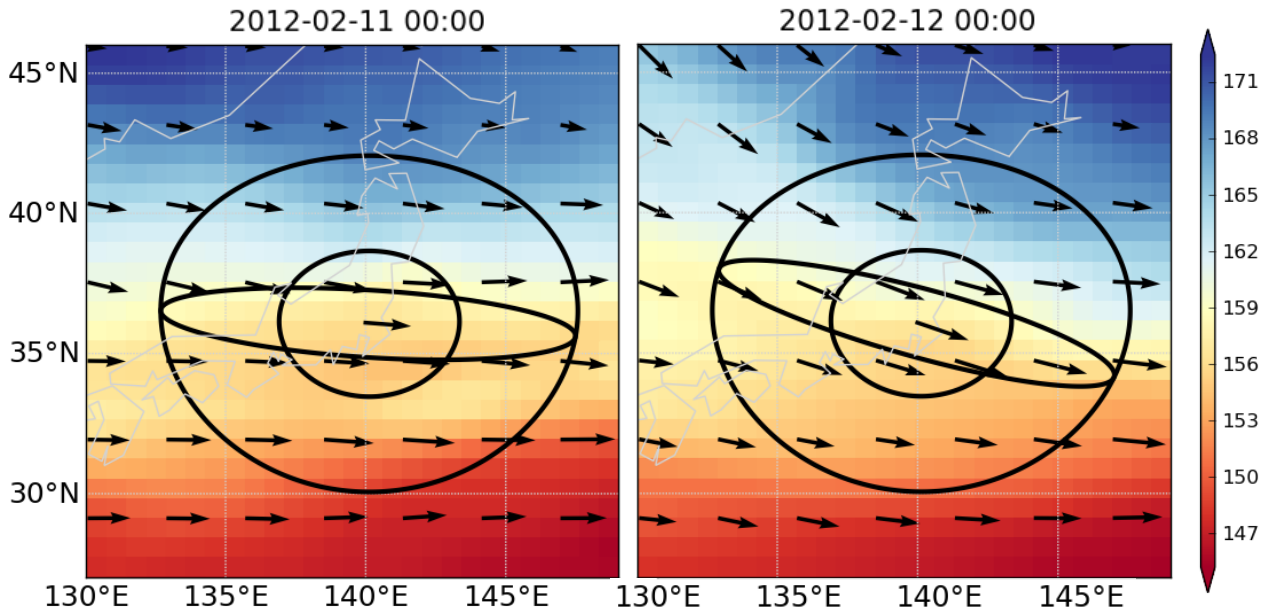
Figure 2. $500\,\mathrm{hPa}$ ERA-Interim refractivity field (color contours) centered at Tateno for Feb. 11 and 12, 2012 at 00:00 UTC. The three spatial geometries (large circle, small circle, ellipse) are shown in black centered at Tateno with wind vectors (black) denoting wind direction and magnitude. Generally, refractivity isopleths lie parallel to the wind flow, even as wind direction changes, resulting in minimal refractivity variability within the ellipse relative to both circles.

$a = b =$ radius, and both a, b are in km):

$$X(s) = a\cos(s)\cos(\theta) - b\sin(s)\sin(\theta)$$
$$Y(s) = a\cos(s)\sin(\theta) + b\sin(s)\cos(\theta) \tag{2}$$

where s is a parameter that varies between 0 and 2π and is stepped in increments of 0.01, yielding a series of 629 points that
5 approximate the circle or ellipse, and θ is the wind direction in radians (converted from meteorological to polar coordinates). The X and Y coordinates of the circle or ellipse are first computed at the Equator where $1\,^\circ$ of latitude and longitude equals $111\,\mathrm{km}$, and then the X and Y coordinates are adjusted to the latitude and longitude of the RS station according to the following:

$$\mathrm{LAT}(Y, s) = \frac{Y(s)}{111} + \mathrm{LAT(RS)}$$
10 $$\mathrm{LON}(X, s) = \frac{X(s)}{111\cos(\mathrm{LAT}(Y, s))} + \mathrm{LON(RS)} \tag{3}$$

The ellipses change orientation such that the semi-major axis is parallel to the wind direction at each pressure level and time per RS. The circles remain fixed and unaffected by the change in wind direction, pressure level, or time.

Figure 2 illustrates the three geometries centered at Tateno at 500 hPa for two days in February 2012, noting that the ellipse adjusts its orientation as wind direction changes with time and at a pressure level. Preliminary testing of the ellipse co-location method with ERA-Interim model refractivity fields confirmed that refractivity isopleths tend to follow wind flow (as illustrated in Fig. 2); thus orientation of the ellipse along the direction of the wind flow should increase atmospheric homogeneity relative to the large circle.

For each RS at a given time and pressure level, RO profiles are co-located with the RS profiles under the time and geometric constraints discussed above. There can (and often are) multiple co-location pairs with the same RS at a given time and pressure level. We computed differences for each co-location pair:

$$X_{\text{RO–RS}} = X_{\text{RO}} - X_{\text{RS}} \tag{4}$$

where X refers to ~~time,~~ refractivity, temperature, ~~and~~ or water vapor pressure. We computed the RMS of the differences for each pressure level over the full time period and used the RMS to quantify the reduction in sampling errors. In certain cases we also computed the percent difference between two quantities X_i and X_j:

$$\% \, diff = \underline{100*}\,100\times\frac{X_i - X_j}{X_j} \tag{5}$$

2.4 ~~Spatial~~ Spatial-temporal sampling correction algorithm
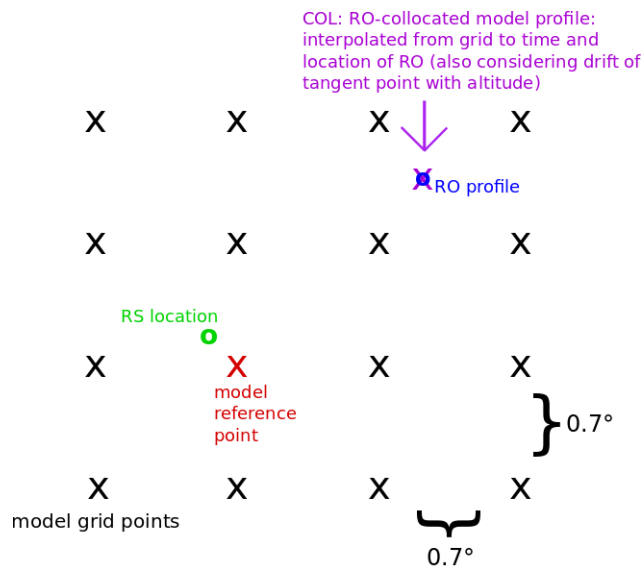
Figure 3. Schematic of the ~~spatial~~ sampling correction applied to RO and RS co-location.

Applying a ~~spatial~~ spatial-temporal sampling correction to the ~~differences during~~ RO–RS ~~co-locations~~ differences is an alternate method of reducing sampling errors in the presence of an auxiliary data set. This method has been applied in previous studies and is not restricted to RO–RS comparisons. Haimberger et al. (2012) used this approach to homogenize RS records. Wong et al. (2015) used ECMWF forecasts ~~to estimate~~ for double-differencing to reduce the sampling differences between Atmospheric Infrared Sounder (AIRS) and RS co-located pairs~~, using a double-difference error estimate to correct AIRS and RS differences~~. Tradowsky et al. (2017) calculated an observed − background (O − B) double-difference to estimate the mean RS temperature bias using co-located RO profiles and ~~the~~ Met Office NWP background fields~~, in which the double-difference is applied to reduce the effects of temporal and spatial sampling errors~~. These studies use a double-difference ~~sampling~~ correction, but do not verify or discuss how the correction reduces sampling errors.

Here, we apply a ~~spatial~~ spatial-temporal sampling correction double-difference computed with model data and assess its effects on reducing sampling errors. We use ERA-Interim data to subtract the model background from both the RO and RS observations, removing spatial and temporal sampling differences and isolating the observational errors associated with the RO and RS pair. As shown in Fig. 3, we use the ~~interpolated values of the~~ ERA-Interim profiles ~~at the RO~~ interpolated in time and space to the RO time and location (eraPrf profiles provided by CDAAC) and ~~use~~ the ERA-Interim grid point nearest the RS location at the RS launch time to compute the ~~spatial~~ sampling correction. Due to the coarser vertical resolution of the ERA-Interim grid relative to the RS and RO vertical resolutions, comparisons with the ~~spatial~~ sampling correction are conducted on a common 50 hPa uniform pressure grid from $1000 - 100$ hPa to avoid further vertical interpolation.

The ~~spatial~~ spatial-temporal sampling corrected differences (X^{sc}) for the co-location pairs are computed as follows for the comparison variables (refractivity, temperature, and water vapor pressure):

$$X^{\text{sc}} = (X_{\text{RO}} - X_{\text{ERA(RO)}}) - (X_{\text{RS}} - X_{\text{ERA(RS)}})$$
$$= (X_{\text{RO}} - X_{\text{RS}}) - (X_{\text{ERA(RO)}} - X_{\text{ERA(RS)}})$$

$$X^{\text{sc}} = (X_{\text{RO}} - X_{\text{ERA(RO)}}) - (X_{\text{RS}} - X_{\text{ERA(RS)}}) \tag{6}$$

$$= (X_{\text{RO}} - X_{\text{RS}}) - (X_{\text{ERA(RO)}} - X_{\text{ERA(RS)}}) \tag{6a}$$

where $X_{\text{ERA(RO)}}$ is the ERA-Interim model value interpolated in time and space to the RO location, $X_{\text{ERA(RS)}}$ is the ERA-Interim model grid point value closet to the RS at the RS launch time, and their difference, $X_{\text{ERA(RO)}} - X_{\text{ERA(RS)}}$, is referred to as the model correction term. We computed the RMS differences per pressure level on the 50 hPa grid.

The spatial-temporal corrected difference (X^{sc}) between the two data sets (Eq. (6) and (6a)) can be interpreted in two ways. When the two data sets are close to each other in time and space, one interpretation is that the RO–RS differences are corrected for local spatial and temporal variability by subtracting these sampling differences, which are estimated by another data set (in this case ERA-Interim) from the measured differences (Eq. (6a)). This is the interpretation used here and by Wong et al. (2015). However, as the equivalent Eq. (6) shows, the corrected difference is also the difference of the departures of the two data sets from a common, reference data set, or the "double-differencing" method (Chander et al., 2013; Tradowsky et al., 2017). In this

interpretation, the two data sets are not necessarily required to be close in space or time; the method is valid as long as the biases of the reference data set do not vary over the spatial and temporal scales of the comparison. We tested the sensitivity of the sampling correction method to the spatial and temporal scale of the comparisons in Appendix A for RO–RS pairs over a much larger spatial scale (within circles of radius 15° latitude) and longer time window (24 h) at Lindenberg. We find that the RMS, mean, and standard deviations of the comparisons are insensitive over this range of spatial and temporal scales, allowing for a large increase in the number of co-located pairs of data.

3   Results and discussion

Preliminary proof-of-concept testing of the ellipse method using only ERA-Interim data demonstrated a significant reduction in RMS refractivity differences within the ellipse relative to both the large circle and circles of similar area to the ellipse (not shown here). In the following two sections, aspects of the ellipse co-location are analyzed at the Lindenberg station. The final section presents the results of both the ellipse co-location and ~~spatial~~ sampling correction at all four RS stations.

3.1   Filtered vs. unfiltered

Table 1. ~~Unfiltered~~ Filtered vs. ~~filtered~~ unfiltered RMS percent differences* for refractivity, temperature, and water vapor pressure at Lindenberg, 2014. RMS differences are computed using all co-located pairs within 100 hPa pressure layers for the ~~unfiltered and~~ filtered and unfiltered data sets, then the percent differences between the two are computed and reported here.

Pressure (hPa)	Refractivity			Temperature			Water Vapor Pressure		
	large circle	small circle	ellipse	large circle	small circle	ellipse	large circle	small circle	ellipse
100 – 200	-2.94 %	-6.87 %	-5.96 %	-3.32 %	-7.99 %	-6.55 %	-0.68 %	-0.56 %	0.50 %
200 – 300	-1.28 %	-3.90 %	-2.34 %	-1.30 %	-4.12 %	-2.46 %	-0.78 %	-1.39 %	-1.65 %
300 – 400	-0.34 %	-0.90 %	-1.03 %	-0.20 %	-0.81 %	-0.71 %	-0.60 %	-0.53 %	-1.02 %
400 – 500	-0.80 %	-2.00 %	-1.22 %	-0.22 %	-0.76 %	-0.49 %	-1.03 %	-2.48 %	-1.32 %
500 – 600	-0.80 %	-1.87 %	-1.23 %	-0.30 %	-0.64 %	-0.49 %	-0.83 %	-2.17 %	-1.38 %
600 – 700	-1.05 %	-1.53 %	-1.36 %	-0.33 %	-0.65 %	-0.65 %	-1.03 %	-1.74 %	-1.59 %
700 – 800	-0.89 %	-1.51 %	-1.21 %	-0.26 %	-0.61 %	-0.65 %	-0.84 %	-1.54 %	-1.18 %
800 – 900	-0.97 %	-1.54 %	-1.12 %	-0.32 %	-0.57 %	-0.21 %	-0.85 %	-1.44 %	-0.99 %
900 – 1000	-0.62 %	-0.65 %	-0.94 %	-0.37 %	-0.41 %	-0.26 %	-0.48 %	-0.61 %	-0.79 %

*percent differences computed as $100 \times (\text{RMS}_{\text{filt}} - \text{RMS}_{\text{unfilt}}) / \text{RMS}_{\text{unfilt}}$

Filtering both the RO and RS profiles has a small, positive impact on reducing RMS differences in refractivity, temperature, and water vapor pressure. Compared to RMS differences computed using the unfiltered profiles at Lindenberg, filtering both the RO and RS profiles before co-location reduces RMS differences by about 1 % on average, up to almost 8 % in some instances (see Table 1). Within the large circle, filtering has less of an impact on reducing vertical representativeness percent errors since sampling errors tend to dominate (mostly due to large spatial differences). The RO–RS percent differences in the small circle

and ellipse are more affected by vertical representativeness errors since the sampling errors are relatively small. Based on these results, ~~to reduce representativeness errors~~ we filtered all RO and RS profiles before computing their differences to reduce representativeness errors.

3.2 Effects of wind speed on ellipse co-location
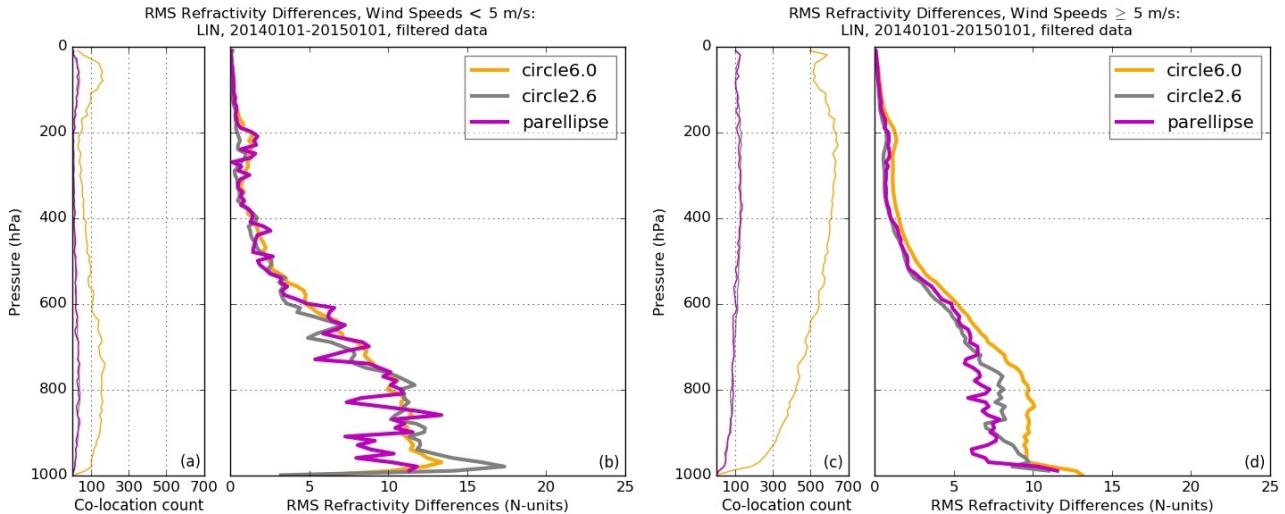
Figure 4. RMS refractivity differences at Lindenberg 2014 for two sets of RO–RS pairs separated by reported RS wind speeds: pairs with wind speeds less than $5\,\mathrm{m\,s^{-1}}$ (left) and pairs with ~~RS~~ wind speeds greater than or equal to $5\,\mathrm{m\,s^{-1}}$ (right). Geometric constraints are shown in the following colors: large circle (orange), small circle (grey), and ellipse (magenta).

5   The relationship between the wind direction and horizontal variability (and sampling error) of refractivity is expected to break down for light wind speeds, and indeed we found that if wind speeds are low, the effectiveness of orienting the ellipse along the direction of wind flow is significantly reduced. We separated co-located RO–RS pairs at Lindenberg into two groups based on the reported wind speed of the RS at a given time and pressure level: (1) wind speeds less than $5\,\mathrm{m\,s^{-1}}$ and (2) wind speeds greater than or equal to $5\,\mathrm{m\,s^{-1}}$. Figure 4 shows the RMS differences in refractivity for each group.

10   As shown in Fig. 4, the clear distinction between RMS profiles for the ellipse and two circles ~~when wind speeds are~~ at wind speeds greater than $5\,\mathrm{m\,s^{-1}}$ essentially vanishes when wind speeds are less than $5\,\mathrm{m\,s^{-1}}$. Similarities in RMS refractivity differences when wind speeds are less than $5\,\mathrm{m\,s^{-1}}$ are likely associated with greater atmospheric homogeneity and weaker gradients in the region, resulting in similar observations within the area containing both circles and the ellipse. When wind speeds increase to larger than $5\,\mathrm{m\,s^{-1}}$, there is a greater separation between the circles and ellipse with respect to the RMS

15   difference. Under moderate to high wind speed conditions, the ellipse reduces RMS refractivity differences relative to both the large and small circle, particularly below about 700 hPa. In the upper troposphere, both the ellipse and the smaller circle give similar results and smaller RMS differences than the large circle (Fig. 4). In the case of Lindenberg, since the majority of wind

speeds are greater than $5 \, \mathrm{m \, s^{-1}}$ (as noted by the differences in co-location counts under the two wind constraints in Fig. 4), low wind speeds do not have a significant effect on the overall RMS differences when co-location pairs are not separated by wind speed (see Fig. 5).

3.3 Ellipse co-location and ~~spatial~~ sampling correction applied at four RS stations
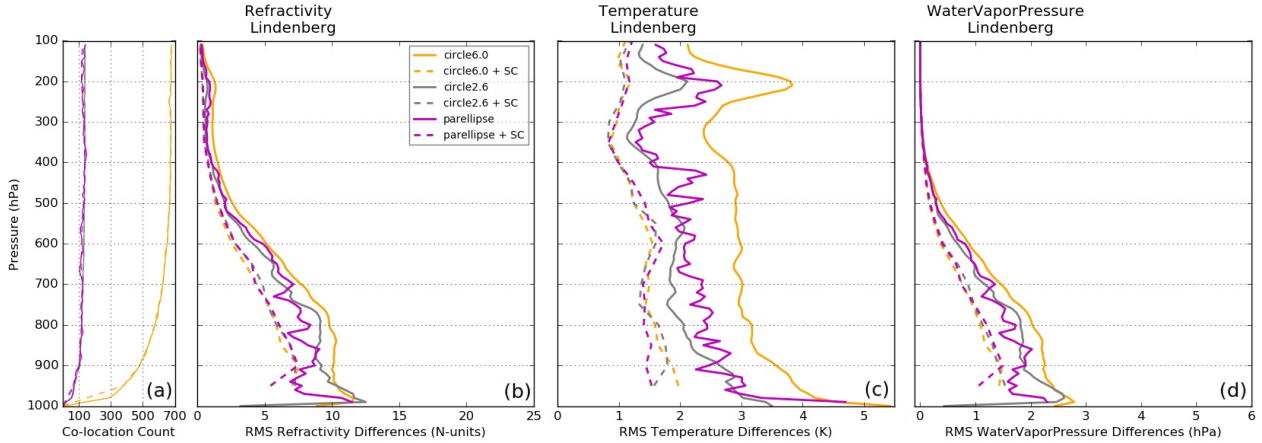
Figure 5. RMS differences for RO–RS co-locations at Lindenberg, 2014. RMS differences with ellipse method only (solid; large circle: orange, small circle: grey, ellipse: magenta) and RMS differences with ellipse method and ~~spatial~~ sampling correction (dashed, abbreviated as SC) are shown for all three variables: (b): refractivity, (c): temperature, (d): water vapor pressure. The number of pairs within each ellipse (solid) and ellipse plus ~~spatial~~ sampling correction (dashed) ~~are~~ is shown in (a).

5   We carried out two RO–RS comparisons at four different RS locations (Lindenberg, Ny-Ålesund, Tateno, and Nauru): first we compared pairs with RO observations co-located within the large circle, small circle, and ellipse centered at the RS station, and second, we applied the ~~spatial~~ sampling correction to the RO–RS pairs within the ellipse and two circles. We then computed RMS differences in refractivity, temperature, and water vapor pressure.

   Figure 5 illustrates the RMS differences for refractivity, temperature, and water vapor pressure with and without the ~~spatial~~
10 sampling correction at Lindenberg. Considering the ellipse method only (solid lines), the ellipse reduces RMS differences relative to the large circle for all three variables at all pressure levels, having the most significant reduction in the temperature RMS differences. Generally, the small circle and ellipse have similar RMS differences, however there are pressure layers in which the ellipse reduces the RMS relative to the small circle. When the ~~spatial~~ sampling correction is applied to the circles and ellipse (dashed lines), the RMS differences are significantly reduced and converge with minimal differences between the
15 ellipse and circles.

   The RMS differences at Lindenberg are mostly affected by spatial sampling errors; temporal sampling errors are minimized by the reduced time window of $1 \, \mathrm{h}$ due to frequent (four times daily) RS launches at this station. ~~The ellipse decreases RMS~~

differences relative to both circles, effectively reducing spatial sampling errors driven by atmospheric variability. The spatial correction is most effective at reducing spatial sampling errors, particularly for the large circle in which atmospheric variability is more likely.
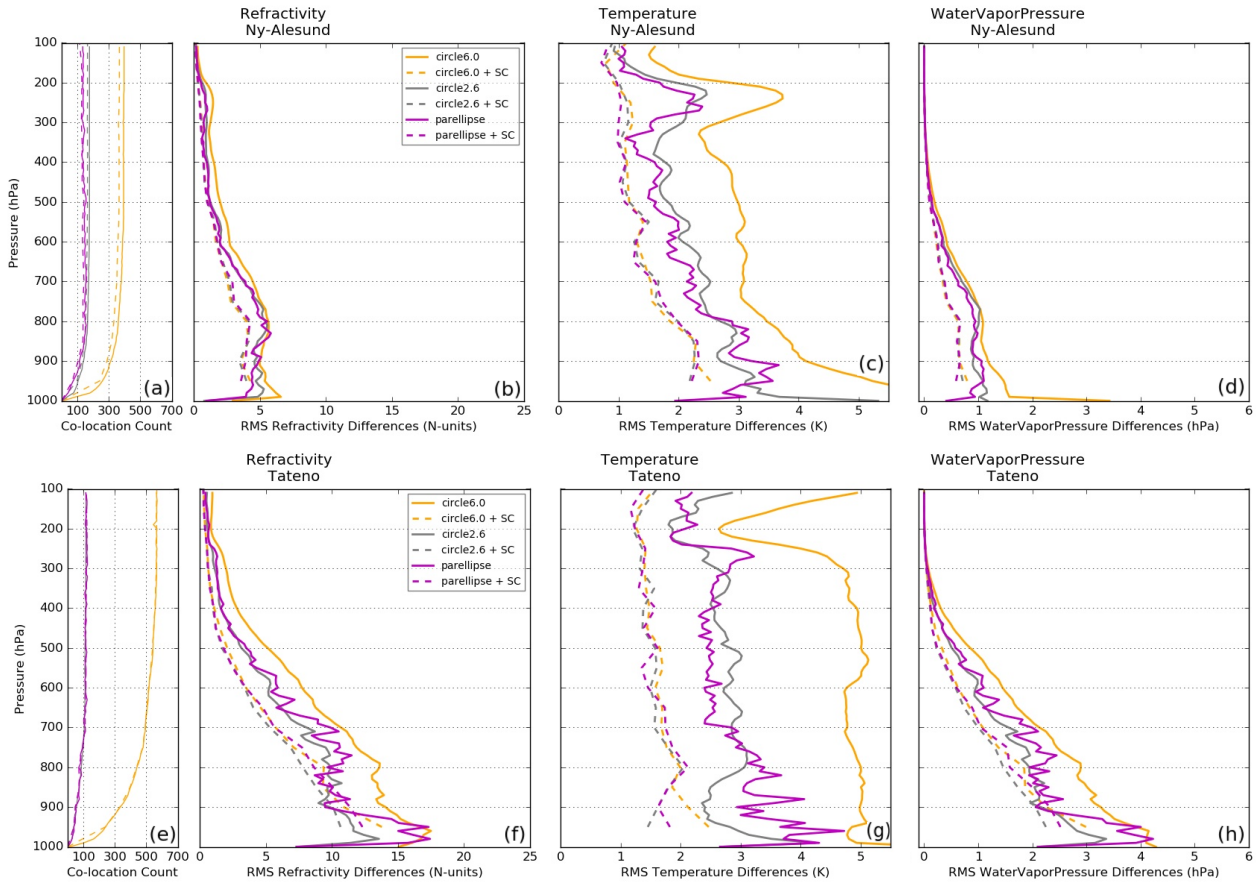
Figure 6. Same as Fig. 5 for two different locations: Ny-Ålesund, 2013 (a-d) and Tateno, 2012 (e-h).

The results at Ny-Ålesund and Tateno are very similar to those at Lindenberg, with some minor differences in the lower

5 troposphere (Fig. 6). Again, using pairs in the ellipse only (solid, magenta) reduces results in significantly reduced RMS differences relative to the large circle (solid, orange) and is equal to or less than RMS differences within smaller RMS differences compared to the small circle (solid, grey). Including the spatial sampling correction shows the largest reduction in RMS differences such that the differences from all three geometric types converge to similar values. At both Ny-Ålesund and Tateno, the reduction of RMS differences in the ellipse compared to the large circle becomes less in the lower troposphere (below

10 $800\,\mathrm{hPa}$) where the frequency of wind speeds less than $5\,\mathrm{m\,s^{-1}}$ increase increases and the relationship between wind direction and RO–RS differences breaks down. For example, at Ny-Ålesund and Tateno the percentages of RO–RS pairs for which

the wind speed is less than $5\,\mathrm{m\,s^{-1}}$ below 800 hPa are 36.1 % and 32.7 % respectively. The small sample size in the ellipse compared to the large circle (Fig. 6) may also play a role, allowing a few outliers to dominate the statistics.
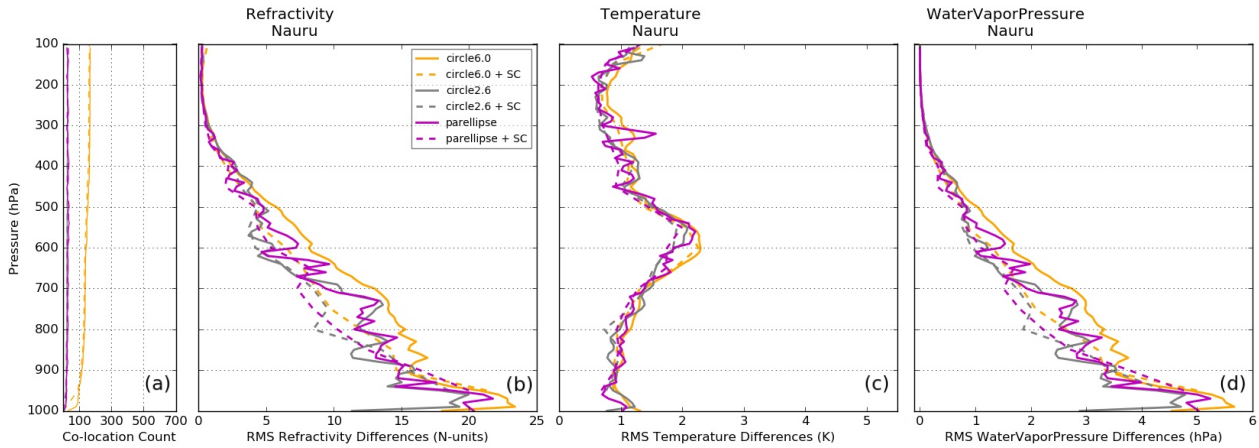
Figure 7. Same as Fig. 5 for Nauru, 2011-2013.

The results at the tropical location of Nauru ~~generally confirms~~ (Fig. 7) generally confirm the findings at the other three RS locations. The ellipse alone (~~Fig. 7,~~ solid, magenta) decreases RMS differences relative to the large circle (solid, orange) and ~~is~~

5   produces RMS differences that are comparable to the ~~RMS~~ differences of the small circle (solid, grey). There is less distinction, however, between the ellipse method only (solid) and ~~spatial~~ sampling correction (dashed) RMS differences, particularly for temperature. All six geometric and ~~spatial~~ sampling correction combinations have overlapping RMS temperature differences and show little separation between co-location methods, unlike the results at Lindenberg, Ny-Ålesund, and Tateno. This is most likely caused by the relative horizontal homogeneity in temperature at Nauru, which is located in the deep tropics. Thus,

10   there is little to no distinction between RMS temperature differences using various geometrical constraints or ~~spatial~~ sampling correction.

Refractivity and water vapor pressure RMS differences in the lower troposphere $(1000 - 700\,\mathrm{hPa})$ at Nauru are the largest of the four stations, caused primarily by atmospheric conditions at its location in the deep tropics. RO is known to have negative refractivity biases in the lower troposphere, particularly in moist tropical regions where large water vapor and associated

15   refractivity gradients often result in super-refraction (Rocken et al., 1997; Ao et al., 2003; Sokolovskiy, 2003; Beyerle et al., 2006; Anthes et al., 2008). Overall, this results in larger RMS refractivity differences at Nauru. Since super-refraction is an error characteristic of RO retrievals and not related to horizontal sampling errors, the ~~spatial~~ sampling correction has no impact on the RO–RS refractivity differences.

4 Conclusions

We have shown that vertical filtering of the RO and RS profiles before comparison reduces representativeness errors associated with different vertical resolutions and observation types by a small amount (typically a few percent). Using these filtered profiles, we tested two methods to reduce spatial and temporal sampling errors during RO–RS ~~co-location~~ comparisons: (1) restricting RO and RS pairs to within ellipses oriented along the direction of wind flow and (2) applying a ~~spatial~~ spatial-temporal sampling correction using model data to remove differences caused by horizontal atmospheric gradients and time differences in the observations. When wind speeds exceed about $5\,\mathrm{m\,s^{-1}}$, co-locations within the ellipse parallel to the wind flow reduce RMS differences in refractivity, temperature, and water vapor pressure relative to co-locations within the large circle, and either reduce or ~~are equal to RMS~~ result in RMS differences that are approximately equal to the differences within the smaller circle. The effectiveness of co-locating RO–RS pairs within the ellipse is reduced for wind speeds less than $5\,\mathrm{m\,s^{-1}}$.

Applying the ~~spatial~~ spatial-temporal sampling correction using ERA-Interim model data showed the most significant reduction in RMS differences, more-so than applying the ellipse constraint alone. The ~~spatial~~ sampling correction reduced RMS differences in refractivity, temperature, and water vapor pressure by an average of $55\,\%$. The reductions of sampling errors within both large and small circles and the ellipse tend to converge with the ~~spatial~~ sampling correction applied, rendering the differences in geometric constraints of the circles and ellipse negligible. An exception to this reduction in RMS occurs at Nauru in the lower troposphere, where super-refraction associated with the atmospheric conditions of the deep tropics tends to dominate the RMS differences.

In order to reduce sampling errors for future RO–RS co-location comparisons, our results suggest that applying the ~~spatial~~ sampling correction under more lenient co-location criteria would be most effective. By using a large distance constraint, the sample size will be sufficiently large and applying the ~~spatial~~ sampling correction eliminates most sampling errors, even for the large distance restriction (greater than $600\,\mathrm{km}$). However, if a reliable model is unavailable, restricting co-locations within ellipses oriented along the direction of wind flow will help to reduce sampling errors caused by atmospheric variability. Both of these methods are effective in reducing sampling errors caused by spatial and temporal differences during comparisons and should provide a more accurate error analysis of RO and RS observations.

Code availability. The code used in this study will be made available upon request.

Data availability. The data used in this study will be made available upon request.

Appendix A

A1 Mean and standard deviation of differences

In this section, we compute the mean and standard deviation (STD) of the RO–RS differences. The mean difference is defined as

$$\mathrm{M(RO-RS)} = \frac{1}{N}\sum_{i=1}^{N}(RO-RS)_i \qquad (A1)$$

The variance about the mean is given by

$$\mathrm{VAR_M(RO-RS)} = \frac{1}{N}\sum_{i=1}^{N}[(RO-RS)_i - M(RO-RS)]^2 \qquad (A2)$$

and the standard deviation of the difference about the mean is the square-root of the variance, $\mathrm{STD_M = VAR_M^{1/2}}$. The $\mathrm{STD_M}$ can be compared to the RMS RO–RS difference, which is defined by

$$\mathrm{RMS(RO-RS)} = [MS(RO-RS)]^{1/2} \qquad (A3)$$

where the mean square difference (MS) is defined as

$$\mathrm{MS(RO-RS)} = \frac{1}{N}\sum_{i=1}^{N}(RO-RS)_i^2 \qquad (A4)$$

It can be shown by expanding Eq. (A2) that

$$\mathrm{VAR_M(RO-RS)} = MS(RO-RS) - [M(RO-RS)]^2 \qquad (A5)$$

Thus, the $\mathrm{STD_M(RO-RS)}$ is always less than or equal to the $\mathrm{RMS(RO-RS)}$, and equal if and only if the mean RO–RS difference is zero.

Figures A1–A3 show the mean and STD of the RO–RS differences for refractivity, temperature, and water vapor pressure, respectively, for all six co-location methods. For each variable, the mean difference profiles are all similar and show no systematic differences, which indicates that the mean differences are not sensitive to the co-location method and are relatively unaffected by sampling errors. This is because the distribution of the RO observations with respect to the RS observations is approximately random, and so the sampling (and representativeness) errors tend to cancel for a large enough sample size. However, the mean differences do illustrate differences in bias errors between the two data sets. The bias errors in the GRUAN RS are expected to be small, so the differences in the mean RO–RS differences are likely due to the RO biases, notably the negative refractivity bias caused by super-refraction in the lower troposphere under moist conditions.

In Fig. A1 we see that the differences in mean refractivity are small above $700\,\mathrm{hPa}$ for all four stations. Below $700\,\mathrm{hPa}$, there is a negative RO refractivity bias, which is most pronounced at the tropical Nauru station where super-refraction is most common. The bias is minimal at the most northern station, Ny-Ålesund, which is expected to have the fewest cases of super-refraction.
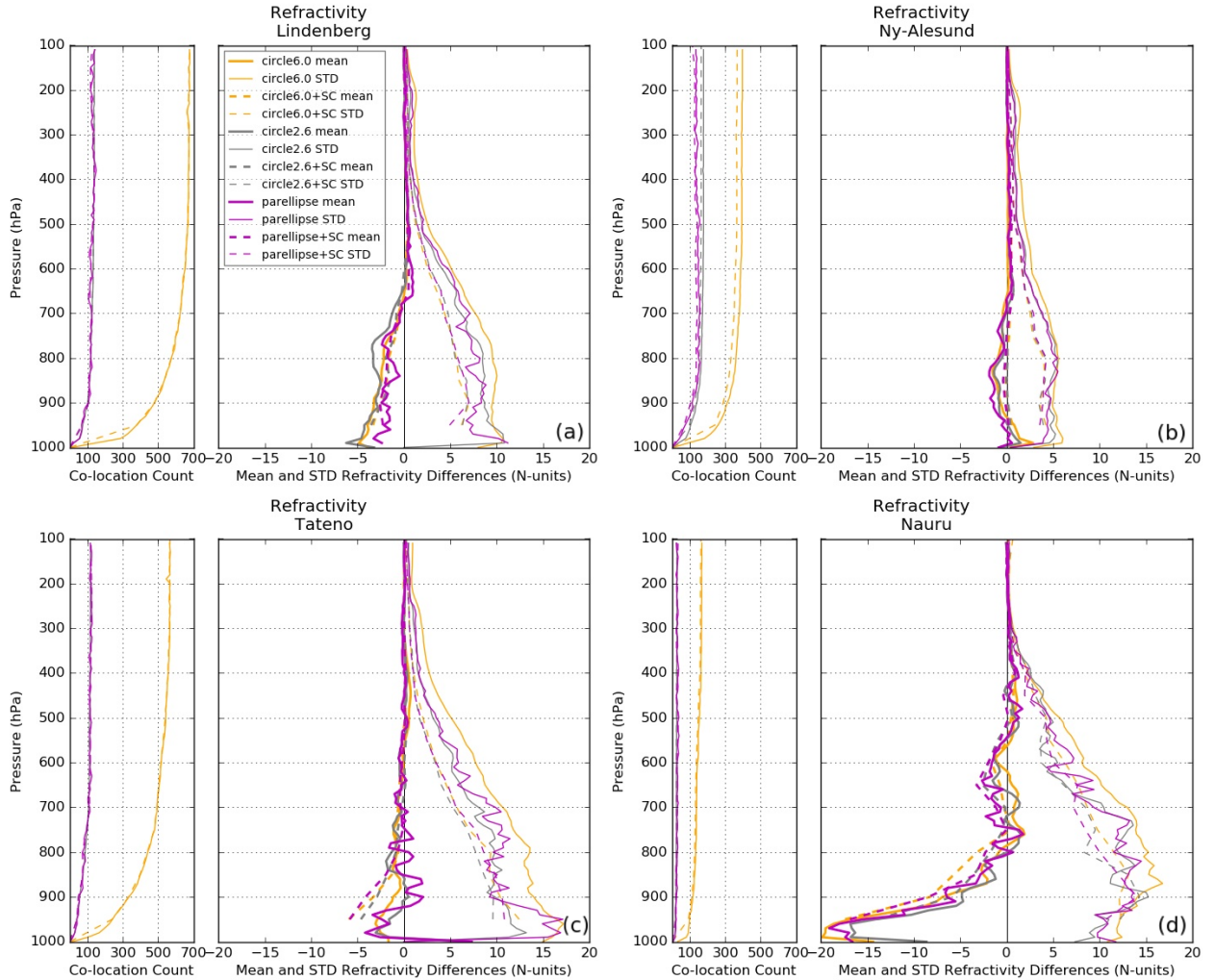
Figure A1. RO–RS refractivity mean (thick) and STD (thin) difference profiles for the six co-location methods at Lindenberg (a), Ny-Ålesund (b), Tateno (c), and Nauru (d). The number of co-located RO–RS pairs are given on the left side of each panel (without the sampling correction: solid; with the sampling correction: dashed). The color schemes are the same as Figs. 4 – 7 and are given in the legend in panel (a) where SC stands for sampling correction.

The biases for temperature differences (Fig. A2) are less than 1 K from 1000 to 100 hPa for Lindenberg, Ny-Ålesund, and Tateno, and show a negative bias at Nauru reaching about 2 K at 600 hPa. This is probably not caused by super-refraction in the RO observations because the negative bias is above the level where most super-refraction occurs and the bias becomes smaller below the 600 hPa level, in contrast to the bias in refractivity which increases sharply with decreasing height in the
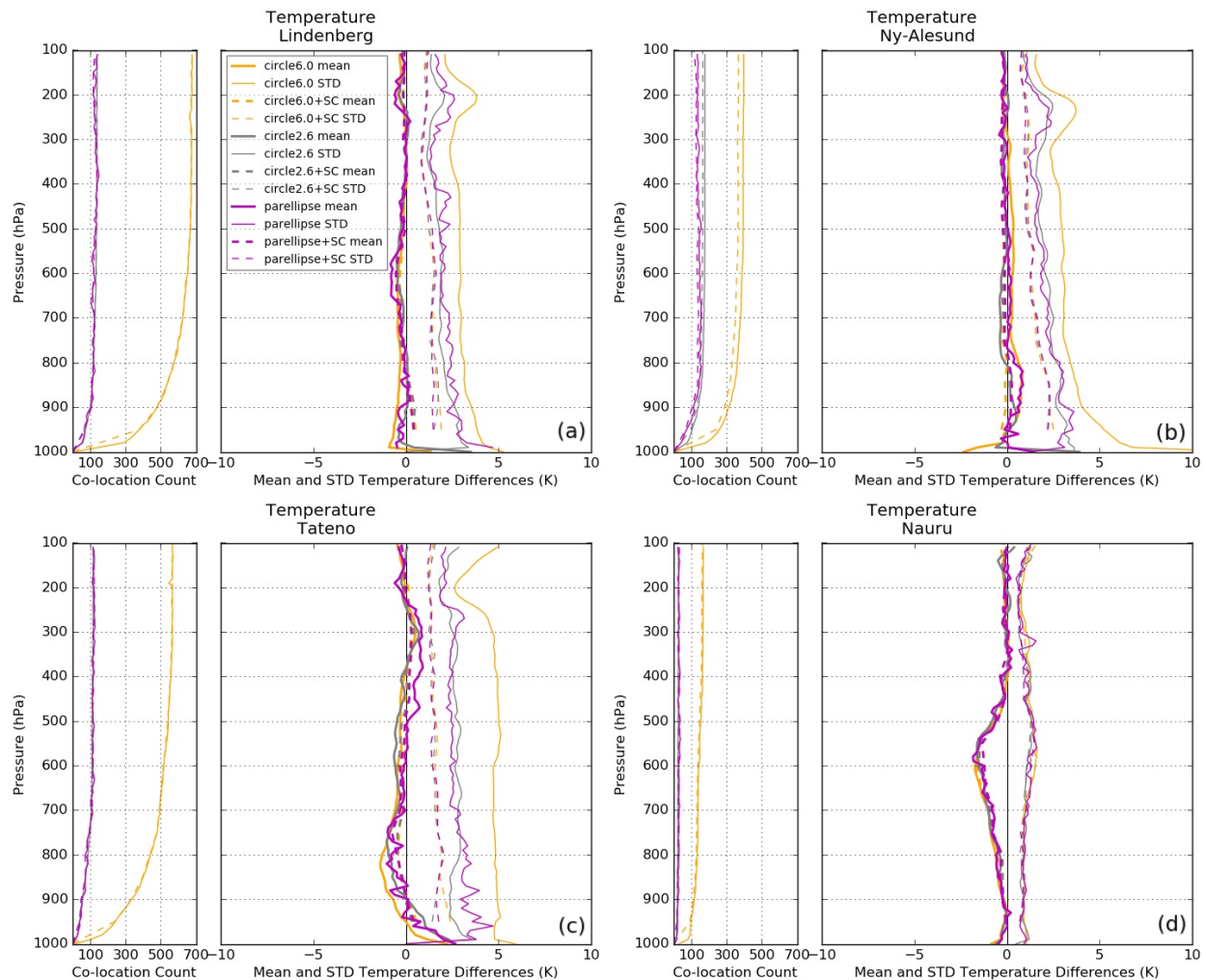
Figure A2. Same as Fig. A1 for temperature.

lower troposphere. Instead, the bias in the temperature is probably related to the 1D-VAR retrieval of temperature in the RO observation, which uses ERA-Interim temperatures as the first guess.

Finally, the differences in the mean RO–RS water vapor pressure reflect the negative refractivity bias due to super-refraction at Lindenberg, Tateno, and especially Nauru, with little bias at the colder, drier Ny-Ålesund station (Fig. A3).

5   The STD profiles in Figs. A1 – A3 are similar to the RMS profiles associated with the different co-location methods in Figs. 5 – 7. This is expected because the STD and RMS values are close when mean RO–RS differences are near zero, as shown by Eq. (A5). The largest differences between the STD and RMS profiles are in the lower troposphere, where the mean differences are the largest.
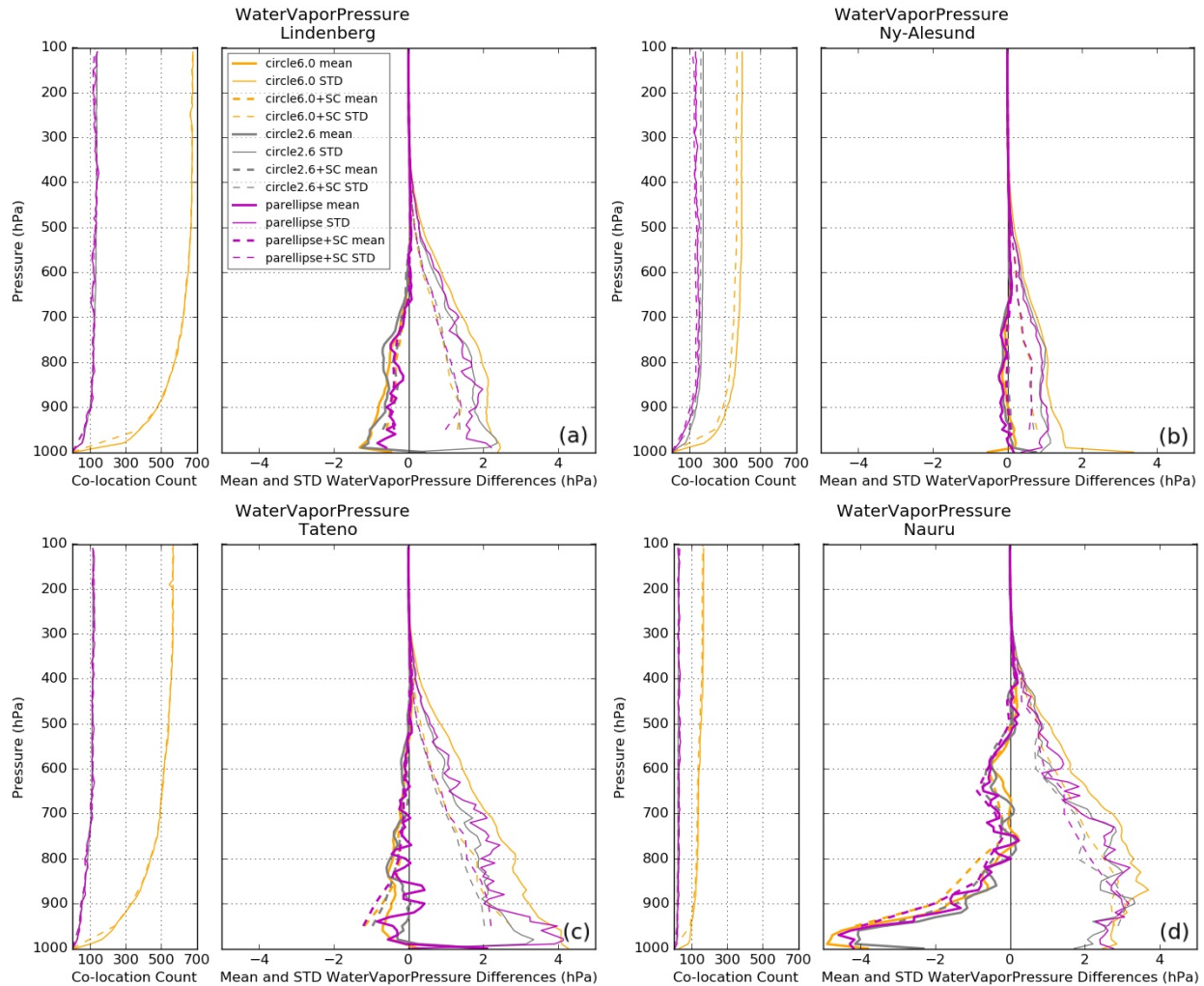
Figure A3. Same as Fig. A1 for water vapor pressure.

A2 Sensitivity of spatial-temporal correction method to spatial and temporal scale

In Sect. 2.4, we noted that when using the double-differencing correction method to reduce spatial and temporal sampling errors, it was not necessary that the RO profiles be close in space and time to the RS profiles, as long as the biases of the reference data set (in our case ERA-Interim) remain constant in space and time throughout the comparison. To illustrate this property of the double-differencing method, we co-located RO profiles within a much larger circle ($15°$ latitude, ~1665 km) and a time window of 24 h at Lindenberg between January and March, 2014, and compared the RMS, mean, and STD of these RO–RS differences to those computed under the $6°$ and $2.6°$ circles with a 1 h time window.
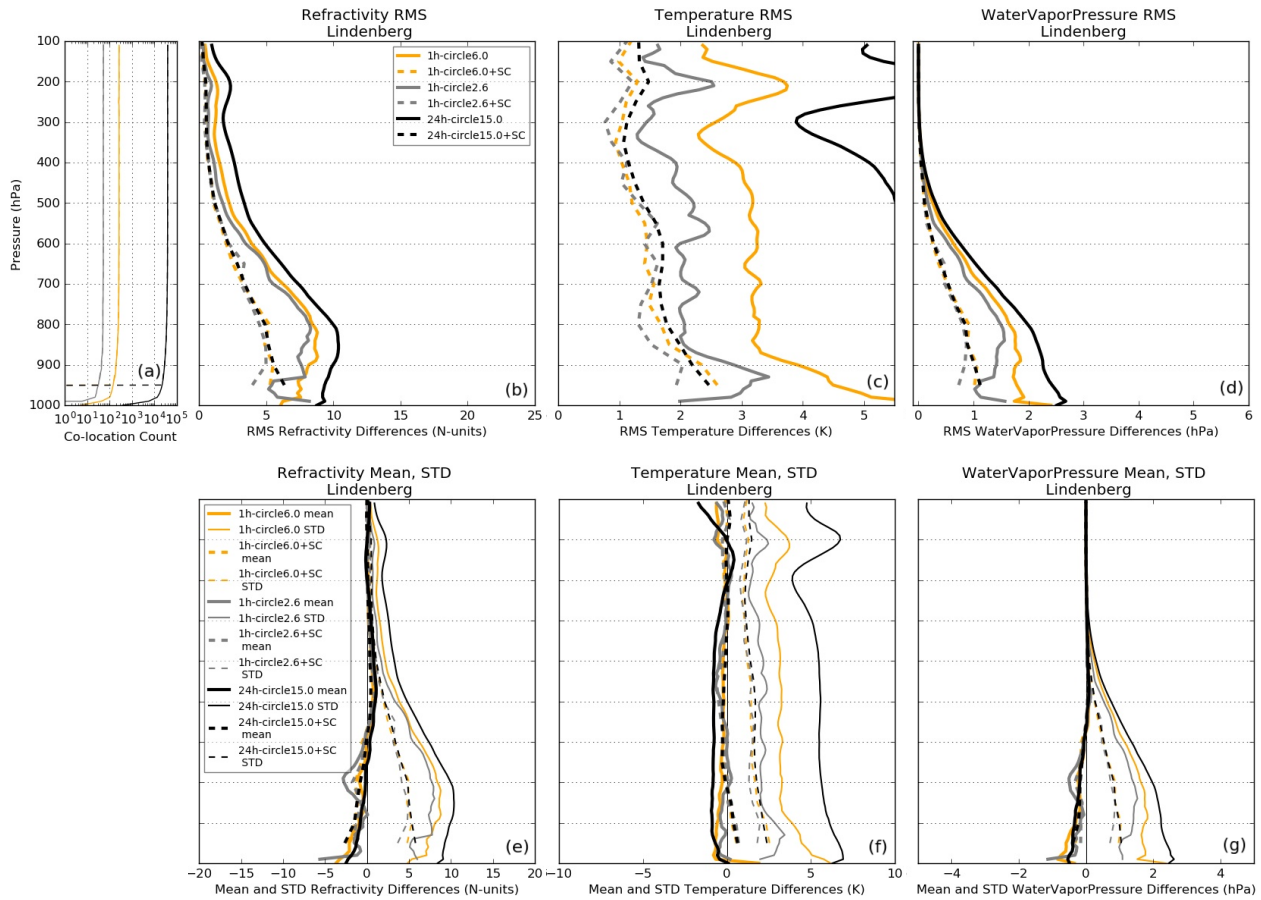
Figure A4. RMS (b-d), mean (thick; e-g), and STD (thin; e-g) RO–RS profiles for refractivity (b, e), temperature (c, f), and water vapor pressure (d, g) under three different co-location restraints with (dashed) and without (solid) the sampling correction (SC) applied: circles of radius 6° latitude (orange) and 2.6° latitude (grey) for time windows of 1 h, and a circle of radius 15° latitude for a 24 h time window (black). Comparisons are conducted at Lindenberg for January – March, 2014. The number of co-located pairs for each co-location restraint (with SC: dashed; without SC: solid) are given in panel (a), where the x-axis is a logarithm scale.

Figure A4 illustrates the RMS (b-d), mean, and STD (e-g) profiles for the RO–RS differences in refractivity, temperature, and water vapor pressure under the three different co-location restraints with and without the sampling correction applied. Across all three variables, the RMS, mean, and STD profiles with the sampling correction remain nearly the same regardless of the spatial or temporal differences applied to the RO–RS co-location. This verifies that the double-differencing method is insensitive to spatial and temporal separations of the RO and RS observations for this example when using ERA-Interim as a reference data set. The similarity in the RMS, mean, and STD profiles with the sampling correction also indicates that any existing bias in the ERA-Interim reference data set is nearly constant over these spatial and temporal scales.

The above results show that using the double-differencing method to reduce spatial and temporal sampling errors in RO–RS comparisons allows for many more RO–RS pairs to be included in the comparison (more than 35,000 for the 15° with 24 h time window compared to approximately 250 for the 6° circle and 1 h time window), as illustrated Fig. A4a, provided that the bias of the reference data set does not vary significantly over the spatial and temporal scales of the comparison. Increasing the size of the spatial and temporal scales of the comparison is thus a tradeoff between reducing the random error effects by increasing sampling size and possibly increasing the errors by allowing a greater effect of varying biases in the reference data set.

Author contributions. All co-authors contributed to developing the ideas and methodologies of this project. SG conducted the majority of the coding, data retrieval, computations, and analysis, with TR contributing to coding and data retrieval. SG prepared the manuscript with contributions from both co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors thank Drs. Eric DeWeaver (NSF) and Jack Kaye (NASA) for their support of this research through NSF-NASA grant AGS-1522830. Dr. Sergey Sokolovskiy provided many useful comments and suggestions during this study. The first author was supported in part by the Significant Opportunities in Atmospheric Research and Science (SOARS) program, NSF grant AGS-1641177, and by the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) program at UCAR, which is sponsored by the National Space Office in Taiwan, NSF, NASA, NOAA, and the U.S. Air Force. Thanks to CDAAC for the provided RO data sets~~and~~, NOAA NCDC and GRUAN for the provided RS data sets.~~ The first and second author would like to ~~, and the ECMWF for the provided ERA-Interim data sets. The authors thank Dr. Keith Maull for his suggestions during this study. The authors would like to thank the three anonymous reviewers for their comments and constructive suggestions which improved this manuscript.

References

Anthes, R., Bernhardt, P., Chen, Y., Cucurull, L., Dymond, K., Ector, D., Healy, S., Ho, S.-P., Hunt, D., Kuo, Y.-H., Liu, H., Manning, K., McCormick, C., Meehan, T., Randel, W., Rocken, C., Schreiner, W., Sokolovskiy, S., Syndergaard, S., Thompson, D. C., Trenberth, K., Wee, T.-K., Yen, N., and Zeng, Z.: The COSMIC/FORMOSAT-3 Mission-Early Results, Bull. Amer. Meteor. Soc., 89, 313–333, https://doi.org/10.1175/BAMS-89-3-31, 2008.

Anthes, R. A.: Exploring Earth's atmosphere with radio occultation: contributions to weather, climate, and space weather, Atmos. Meas. Tech., 4, 1077–1103, https://doi.org/10.5194/amt-4-1077-2011, 2011.

Ao, C., Meehan, T., Hajj, G., Mannucci, A., and Beyerle, G.: Lower troposphere refractivity bias in GPS occultation retrievals, J. Geophys. Res., 108, https://doi.org/10.1029/2002JD003216, 2003.

Beyerle, G., Schmidt, T., Wickert, J., Heise, S., Rothacher, M., Konig-Langlo, G., and Lauristen, K.: Observations and simulations of receiver-induced refractivity biases in GPS radio occultation, J. Geophys. Res., 111, https://doi.org/10.1029/2005JD006673, 2006.

Bodeker, G., Bojinski, S., Cimini, D., Dirksen, R., Haeffelin, M., Hannigan, J., Hurst, D., Leblanc, T., Madonna, F., Maturilli, M., Mikalsen, A., Philipona, R., Reale, T., Seidel, D., Tan, D., Thorne, P., Vömel, H., and Wang, J.: Reference upper-air observations for climate: from concept to realty, Bull. Amer. Meteor. Soc., 97, 123–135, https://doi.org/10.1175/BAMS-D-14-00072.1, 2016.

Bruce, R., Duncan, L., and Pierluissi, J.: Experimental study of the relationship between radiosonde temperatures and satellite-derived temperatures, Mon. Wea. Rev., pp. 493–496, 1977.

Chander, G., Hewison, T., Fox, N., Wu, X., Xiong, X., and Blackwell, W.: Overview of intercalibration of satellie instruments, IEEE Trans. on Geoscience and Remote Sensing, 15, 1056–1080, https://doi.org/10.1109/TGRS.2012.2228654, 2013.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Koehler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quart. J. Roy. Meteor. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

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

Eyre, J.: Assimilation of radio occultation measurements into a numerical weather prediction system, Technical Memorandum 199, European Centre for Medium-Range Weather Forecasts, Reading, UK, 1994.

Fasso, A., Ignaccolo, R., Madonna, F., Demoz, B., and Franco-Villoria, M.: Statistical modelling of collocation uncertainty in atmospheric thermodynamic profiles, Atmos. Meas. Tech., 7, 1803–1816, https://doi.org/10.5194/amt-7-1803-2014, 2014.

Haimberger, L., Tavolato, C., and Sperka, S.: Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighboring stations, J. Climate, 25, 8108–8131, https://doi.org/10.1175/JCLI-D-11-00668.1, 2012.

He, W., Ho, S.-P., Chen, H., Zhou, X., Hunt, D., and Kuo, Y.-H.: Assessment of radiosonde temperature measurements in the upper troposphere and lower stratosphere using COSMIC radio occultation data, Geophys. Res. Lett., 36, https://doi.org/10.1029/2009GL038712, 2009.

Ho, S.-P., Zhou, X., Kuo, Y.-H., Hunt, D., and Wang, J.: Global evaluation of radiosonde water vapor systematic biases using GPS radio occultation from COSMIC and ECMWF analysis, Remote Sensing, 2, 1320–1330, https://doi.org/10.3390/rs2051320, 2010.

Ho, S.-P., Peng, L., and Vömel, H.: Characterization of the long-term radiosonde temperature bias in the upper troposphere and lower strato-
sphere using COSMIC and Metop-A/GRAS data from 2006 to 2014, Atmos. Chem. Phys., 17, 4493–4511, https://doi.org/10.5194/acp-
17-449-2017, 2017.

Ignaccolo, R., Fraco-Villoria, M., and Fasso, A.: Modelling collocation uncertainty of 3D atmospheric profiles, Stoch. Environ. Res Risk
5   Assess, https://doi.org/10.1007/s00477-014-0890-7, 2015.

Immler, F., Dykema, J., Gardiner, T., Whiteman, D., Thorne, P., and Vömel, H.: Reference quality upper-air measurements: guidance for
developing GRUAN data products, Atmos. Meas. Tech., 3, 1217–1231, https://doi.org/10.5194/amt-3-1217-2010, 2010.

Kitchen, M.: Representativeness errors for radiosonde observations, Quart. J. Roy. Meteor. Soc., 115, 673–700, 1989.

Kuo, Y.-H., Wee, T., Sokolovskiy, S., Rocken, C., Schreiner, W., Hunt, D., and Anthes, R.: Inversion and error estimation of GPS radio
10   occultation data, J. Meteor. Soc. Japan, 82, 507–531, https://doi.org/10.2151/jmsj.2004.507, 2004.

Kuo, Y.-H., Schreiner, W., Wang, J., Rossiter, D., and Zhang, Y.: Comparison of GPS radio occultation soundings with radiosondes, Geophys.
Res. Lett., 32, https://doi.org/10.1029/2004GL021443, 2005.

Kursinski, E., Hajj, G., Schofield, J., Linfield, R., and Hardy, K.: Observing Earth's atmosphere with radio occultation measurements using
the Global Positioning System, J. Geophys. Res., 102, 23 429–23 465, https://doi.org/10.1029/97JD01569, 1997.

15 Ladstädter, F., Steiner, A., Schwärz, M., and Kirchengast, G.: Climate intercomparison of GPS radio occultation, RS90/92 radiosondes and
GRUAN from 2002 to 2013, Atmos. Meas. Tech., 8, 1819–1834, https://doi.org/10.5194/amt-8-1819-2015, 2015.

Lohmann, M.: Analysis of Global Positioning System (GPS) radio occultation measurement errors based on Satellite de Aplica-
ciones Cientificas-C (SAC-C) GPS radio occultation data recorded in open-loop and phase-locked-loop mode, J. Geophys. Res., 112,
https://doi.org/10.1029/2006JD007764, 2007.

20 Mears, C., Wang, J., Smith, D., and Wentz, F.: Intercomparison of total precipitable water measurements made by satellite-borne microwave
radiometers and ground-based GPS instruments, J. Geophys. Res., 120, 2492–2504, https://doi.org/10.1002/2014JD022694, 2015.

Melbourne, W., Davis, E., Duncan, C., Hajj, G., Hardy, K., Kursinski, E., Meehan, T., Young, L., and Yunck, T.: The application of space
borne GPS to atmospheric limb sounding and global change monitoring, Tech. rep., Jet Propulsion Lab., California Institute of Technology,
1994.

25 Rocken, C., Anthes, R. A., Exner, M., Hunt, D., Sokolovskiy, S., Ware, R., Gorbunov, M., Schreiner, W., Feng, D., Herman, B., Kuo,
Y.-H., and Zou, X.: Analysis and validation of GPS/MET data in the neutral atmosphere, J. Geophys. Res., 102, 29 849–29 866,
https://doi.org/10.1029/97JD02400, 1997.

Savitzky, A. and Golay, M.: Smoothing and differentiation of data by simplified least squares procedure, Analytical Chemistry, 36, 1627–
1639, 1964.

30 Seidel, D., Sun, B., Pettey, M., and Reale, A.: Global radiosonde balloon drift statistics, J. Geophys. Res.,
https://doi.org/10.1029/2010JD014891, 2011.

Seidel, D. J., Berger, F. H., Diamond, H. J., Dykema, J., Goodrich, D., Immler, F., Murray, W., Peterson, T., Sisterson, D., Sommer, M.,
Thorne, P., Voemel, H., and Wang, J.: REFERENCE UPPER-AIR OBSERVATIONS FOR CLIMATE Rationale, Progress, and Plans,
Bull. Amer. Meteor. Soc., 90, 361–369, https://doi.org/10.1175/2008BAMS2540.1, 2009.

35 Smith, E. K. and Weintraub, S.: The constants in the equation for atmospheric refractive index at radio frequencies, Proceedings of the IRE,
41, 1035–1037, 1953.

Sokolovskiy, S.: Effect of superrefraction on inversions of radio occultation signals in the lower troposphere, Radio Sci., 38,
https://doi.org/10.1029/2002RS002728, 2003.

Staten, P. and Reichler, T.: Apparent precision of GPS radio occultation temperatures, Geophys. Res. Lett., 36, https://doi.org/10.1029/2009GL041046, 2009.

Sun, B., Reale, A., Seidel, D. J., and Hunt, D. C.: Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics, J. Geophys. Res., 115, https://doi.org/10.1029/2010JD014457, 2010.

Tradowsky, J., Burrows, C., Healy, S., and Eyre, J.: A new method to correct radiosonde temperature biases using radio occultation data, J. Appl. Meteor. Climatol., 56, 1643–1661, https://doi.org/10.1175/JAMC-D-16-0136.1, 2017.

Vergados, P., Mannucci, A., and Ao, C.: Assessing the performance of GPS radio occultation measurements in retrieving tropospheric humidity in cloudiness: A comparison study with radiosondes, ERA-Interim, and AIRS data sets, J. Geophys. Res., 19, 7718–7731, https://doi.org/10.1002/2013JD021398, 2014.

Wang, B.-R., Liu, X., and Wang, J.-K.: Assessment of COSMIC radio occultation retrieval products using global radiosonde data, Atmos. Meas. Tech., 6, 1073–2013, https://doi.org/10.5194/amt-6-1073-2013, 2013.

Ware, R., Exner, M., Feng, D., Gorbunov, M., Hardy, K., Herman, B., Kuo, Y.-H., Meehan, T., Melbourne, W., Rocken, C., Schreiner, W., Sokolovskiy, S., Solheim, F., Zou, X., Anthes, R. A., Businger, S., and Trenberth, K.: GPS sounding of the atmosphere from low earth orbit: Preliminary results, Bull. Amer. Meteor. Soc., 77, 19–40, https://doi.org/10.1175/1520-0477(1996)077<0019:GSOTAF>2.0.CO;2, 1996.

Wickert, J., Schmidt, T., Beyerle, G., Konig, R., and Reigber, C.: The radio occultation experiment aboard CHAMP: Operational data analysis and validation of vertical atmospheric profiles, J. Meteor. Soc. Japan, 82, 381–395, https://doi.org/10.2151/jmsj.2004.381, 2004.

Wong, S., Fetzer, E., Schreier, M., Manipon, G., Fishbein, E., Kahn, B., Yue, Q., and Irion, F.: Cloud-induced uncertainties in AIRS and ECMWF temperature and specific humidity, J. Geophys. Res., 120, 1880–1901, https://doi.org/10.1002/2014JD022440, 2015.

Xu, X., Luo, J., and Shi, C.: Comparison of COSMIC radio occultation refractivity profiles with radiosonde measurements, Adv. Atmos. Sci., 26, 1137–1145, https://doi.org/10.1007/s00376-009-8066-y, 2009.

Zhang, K., Fu, E., Silcock, D., Wang, Y., and Kuleshov, Y.: An investigation of atmospheric temperature profiles in the Australian region using collocated GPS radio occultation and radiosonde data, Atmos. Meas. Tech., 4, 2087–2092, https://doi.org/10.5194/amt-4-2087-2011, 2011.