We thank the three referees for their thoughtful comments. We have revised the paper after carefully considering their comments. Their comments (in italics) and our responses are addressed below. 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.

Reviewer 1: Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-487, 2018

General Comments:
The authors estimate the observation and model error variances by using the "N-cornered hat method". In this study, they estimate the error variances for observation and model in several variables, e.g., refractivity, temperature, specific humidity and relative humidity. They compare their results with previous studies and find that the error patterns are consistent. The errors characters for the GPS RO retrieved temperature and moisture are rarely discussed in previous studies, and it is good to see the estimation in the study. For the manuscript, I have some comments as follows.

Specific comments:
1. The manuscript discusses the observational error variances for the refractivity, temperature, and moisture (q and RH) from GPS RO, but not for the bending angle. The bending angle from GPSRO has been assimilated in several operational centers for weather forecast. Is it possible to provide the error estimation for the bending angle as well? This could be interesting and useful for community users and/or the NWP people.

We agree that the bending angle (BA) error variance estimates would be very interesting. However, as shown by Gilpin et al. (2018b), adding BA to the observations currently in the study (refractivity, temperature, relative humidity and specific humidity) is not trivial. To use the 3CH method, BA corresponding to radiosonde, ERA-Interim and GFS data sets would have to be computed from the refractivity of these data sets using a forward model (the Abel Integral). The Abel Integral requires vertical differentiation of refractivity, which requires accurate vertical interpolation of refractivity between the model (or radiosonde) levels. The vertical resolutions of the models (and often also the radiosonde) are low compared to the resolution required to accurately approximate the Abel Integral. For these low-resolution data sets, small-scale atmospheric structures as well as the quasi-exponential variation of refractivity with height, can create large errors in the calculated bending angle. Thus, to add discussion on the forward model and interpolation methods and then to compute error variances for bending angles is outside the scope of this study.

2. The samples are picked up within 600km and 3h for comparison, is the criteria the same for ERA and GFS? For example, do the authors apply a spatiotemporal interpolation when comparing ERA and RO? If it uses the co-location criteria, how much could this affect the error variance?

In response to this comment, we have added more detail to how the data sets are co-located. We have added a new Section 2.5 (on Page 4) titled "Co-location of the data sets":

"The locations of the four radiosonde stations are chosen for the comparisons. We use RO observations that are located within 600 km and 3 hours of the radiosonde launches. CDAAC provides GFS and ERA profiles that are already linearly interpolated in space and time to the RO location and time. These interpolated profiles, along with the RO observations, were corrected for their time and spatial differences from the radiosonde data using a model correction algorithm (Gilpin et al., 2018). Thus the effect of spatial and temporal differences among the data sets is expected to be minor."

3. The abstract does not point out the major conclusion of the study. I would suggest to add the part into the abstract.

We agree and have added the following paragraph to the abstract:

“Our results show that different combinations of the four data sets yield similar relative and specific humidity, temperature, and refractivity error variance profiles at the four stations, and these estimates are consistent with previous estimates where available. These results thus indicate that the correlations of the errors among all data sets are small and the 3CH method yields realistic error variance profiles. The estimated error variances of the ERA-Interim data set are smallest, a reasonable result considering the excellent model and data assimilation system and assimilation of high-quality observations. For the four locations studied, RO has smaller error variances than radiosondes, in agreement with previous studies.”

4. The statistics are based on samplings near the four stations. According to previous studies, the observational error variance could vary with latitudes, can the results in this study be applied globally?

The main point of this paper is to show how the 3CH method can be used to estimate vertical profiles of temperature, relative and specific humidity, and refractivity error variances of five data sets at selected challenging locations in the tropics and sub-tropics. The error variances of all data sets are likely to vary with latitude, and also over different regions and seasons (at least for the models and radiosondes with their different radiosonde sensor types). A study using different locations, seasons, and years would be interesting and would be a better indicator of the error variances on a global scale. A calculation of these error variances at many locations over a number of years would also be interesting, to see how the errors of the various systems vary over time.

5. In Fig. 5a, A1 and A3, there are several gaps in STD (ERA-True). The ERA-Interim data should be continuous.

The ERA data themselves are continuous, but the estimated ERA-Interim error variances are close to zero. Furthermore, we neglect the error covariance terms in the computation and have a limited sample size. The combination of these three factors can lead to negative estimated error variances at times when the true error variance is so close to zero, and the STD is thus undefined. This is discussed briefly at the end of Section 3 and also in Appendix A, with references.

We added the following sentence on page 9 lines 5-7:
“The gap in the computed ERA error STD in Fig. 5a occurs due to negative estimated error variance values, which can result from having a limited sample size, neglecting error covariance terms during computation, and having an error variance that is already close to zero (as is the case for ERA).”

6. In the manuscript, the "N" can be represented for several meanings, e.g., RO refractivity, the number of samples for statistic, and number of data sets, etc. It would be better to use different characters to avoid confusing.

Thank you. We now use "*N*" for refractivity and "*n*" for the number of samples now throughout the manuscript.

7. Page 4 line 15: There are three types of COSMIC data provided from CDAAC, i.e., re-processed, post-processed, and real-time data. In this manuscript, it uses the ERA-Interim data as the background for the 1DVAR retrieval, do the authors get the COSMIC data from re-processed data?

We use only COSMIC re-processed data in this study. We added this information in Section 2.3 describing RO.

Technical corrections:
1. The variables should be in the same form with italics, for example, page 4 line 17: "Specific humidity q*" → please change to "Specific humidity q"; page 4 line 18: "water vapor pressure* e*" → "water vapor pressure e".*

Corrected

2. Page 22 line 14 and line 16: three-cornered hat (THC) → three-cornered hat (TCH)

Corrected to 3CH, which we now use throughout the paper.

3. Pages 24 and 25: The descriptions for equations (13)-(15) are inconsistent, for example, the equation (13) is computed from Eqs. (3), (5) and (6), not (1), (2) and (3) that indicated in line 26.

We have eliminated the derivation and associated discussion of the linearly dependent equations for estimating the error variances and so these equations no longer exist.

4. Page 29 Fig. A4 figure caption: Mean of six...
of normalized "specific humidity" → should be "refractivity"

Thank you. We have deleted this figure from the revised paper.

Reviewer 2: Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-487, 2018 Received and published: 19 April 2018

This manuscript applies the N-cornered hat technique to estimate errors in geophysical measurements: radio occultation, radiosondes, ERA-Interim reanalysis and weather forecast outputs at four locations in the tropics and subtropics. The N-cornered hat technique is closely related to the method of triple collocation, which has been widely applied to geophysical datasets in the literature (as the authors note). However, some subtle differences between the approaches are missed in this analysis, which may impact the results.

One important difference between triple collocation (TC) and the three-cornered hat (3CH) is the treatment of the underlying truth. TC treats the underlying truth as a random

variable (Stoffelen 1998), whereas 3CH does not. As a result, TC requires an additional assumption compared to 3CH: the errors must be uncorrelated with the underlying truth. Since the underlying truth is not considered to be a random variable in 3CH, the correlation between the errors and truth is always zero. So, should the underlying truth be treated as a random variable (as in TC) or as deterministic (as in 3CH)? I would argue that, for assessing the stability of clocks, the assumption of a deterministic underlying truth is quite reasonable. However, when considering atmospheric applications, as in this study, it is hard to justify. The atmosphere is a chaotic system with substantial internal variability. Differences between measurements can be due to measurement errors, but they can also be due to the internal variability of the chaotic system. In Figure 3a, for example, there are clear differences between the specific humidity profile estimated by GFS and RS on 13 January, 2007 at 00:23 UTC. Even if the measurement and modelling errors of both GFS and RS were both zero, we would expect there to be some difference between these two profiles because of the internal variability of the system, even accounting for some assimilation of observations into the GFS. Yet the 3CH implicitly attributes ALL differences between different measurement types to measurement errors in one or more measurement types. This seems misguided and is one of the reasons TC is typically applied to geophysical measurements rather than 3CH. Treating the underlying truth as deterministic rather than random leads the authors to neglect the impacts of possible covariance between the errors and the underlying truth, which is likely biasing the error estimates in this study.

The 3CH method allows for a temporally varying "truth" and is appropriate for complex geophysical systems such as the atmosphere. For example, it has been successfully used by Valty et al. (2013) to estimate the geophysical load deformation computed from GRACE satellites, GPS vertical displacement measurements, and global general circulation (GCM) models. O'Carroll et al. (2008) estimated the errors associated with two radiometer and buoy measurements of sea-surface temperatures using exactly the same equations we used. They also provide an excellent discussion on "truth," and in particular conclude that the basic equations in the 3CH method (our Eqs. (7)-(9)) "continues to hold regardless of whatever our definition of the true value might be."

The differences associated with the "internal variability" of the atmosphere as illustrated in Fig. 3a are partly due to measurement errors and partly due to representativeness errors, as the reviewer notes. Both types of errors are included in the error estimates using the 3CH methods.

The main difference between the 3CH and TC method is that the TC method corrects for additive and multiplicative biases among the three data sets, as discussed by Stoffelen (1998), Vogelzang et al. (2011) and others. The TC method calibrates two of the data sets against the third, eliminating biases among the three data sets.

We have added a new section (A3) in Appendix A that compares the 3CH and TC method for a subset of our data sets. Also, Appendix A now includes more discussion of the differences between the two methods and the additive and multiplicative biases. The results in A3 of the Appendix show that the two methods give very similar results (please see next response).

A second important difference is that TC accounts for multiplicative biases in the measurements in a way that 3CH does not. 3CH implies the following measurement model:
$i = T + e_i$ . . . (1)
where T is the unknown true value, and e_i is the error in measurement i. In contrast, triple collocation starts with the measurement model:
$i = a_i + b_i*T + e_i$ . . . (2)
where a_i and b_i are additive and multiplicative biases, respectively (see Gruber et al. (2016),

equation 1). The advantage of the measurement model in TC is that it is robust to multiplicative biases. If multiplicative biases exist but are not factored into the measurement model (as in 3CH), the multiplicative biases will lead to substantial correlations between the errors and the unknown truth. In turn, the neglect of these correlations will bias 3CH error estimates.

Given both of these problems with the analysis, it is not surprising that the estimated errors (for example, in Figure 10) are not internally consistent (as they should be, if either technique – 3CH or TC – were applied appropriately).

We have applied the 3CH method appropriately. Our equations are correct and all assumptions are stated clearly. As mentioned above, the 3CH method has been applied successfully to other geophysical data sets. Biases in the measurements may be present, but tend to cancel as discussed in our discussion paper AMT-2018-75 "Evaluating two methods of estimating error variances from multiple data sets using an error model."

We do not consider the results shown in Fig. 10 to be inconsistent. Using different combinations of data sets to estimate the error variance profiles will lead to different estimates because of different (though small) error covariances in the different data sets and the limited sample size. This is also shown using the error model and idealized data sets in AMT-2018-75. We describe this in the summary and give more detail in Appendix A where the variation of three estimates is discussed: "If all the neglected error covariance terms were in fact identically zero and the sample size was very large (much larger than our sample size), all three estimates of the error variances would be the same. The fact that they give different solutions is because the neglected COV_{err} terms are in reality not zero, and hence their neglect affects the three approximate equations in different ways to give three different solutions. The relatively small sample size *n* also contributes to the differences in the three solutions, which are a measure of these effects."

Furthermore, as discussed above, we computed the estimated error variances for specific humidity using the TC method (including multiplicative and additive biases) and showed that the results are very similar to the 3CH method, confirming that the effect of biases is small in the 3CH method for these data sets. Please see two figures below for radio occultation (RO), radiosonde (RS), ERA, and GFS. We have added these results to Appendix A in the revised paper.

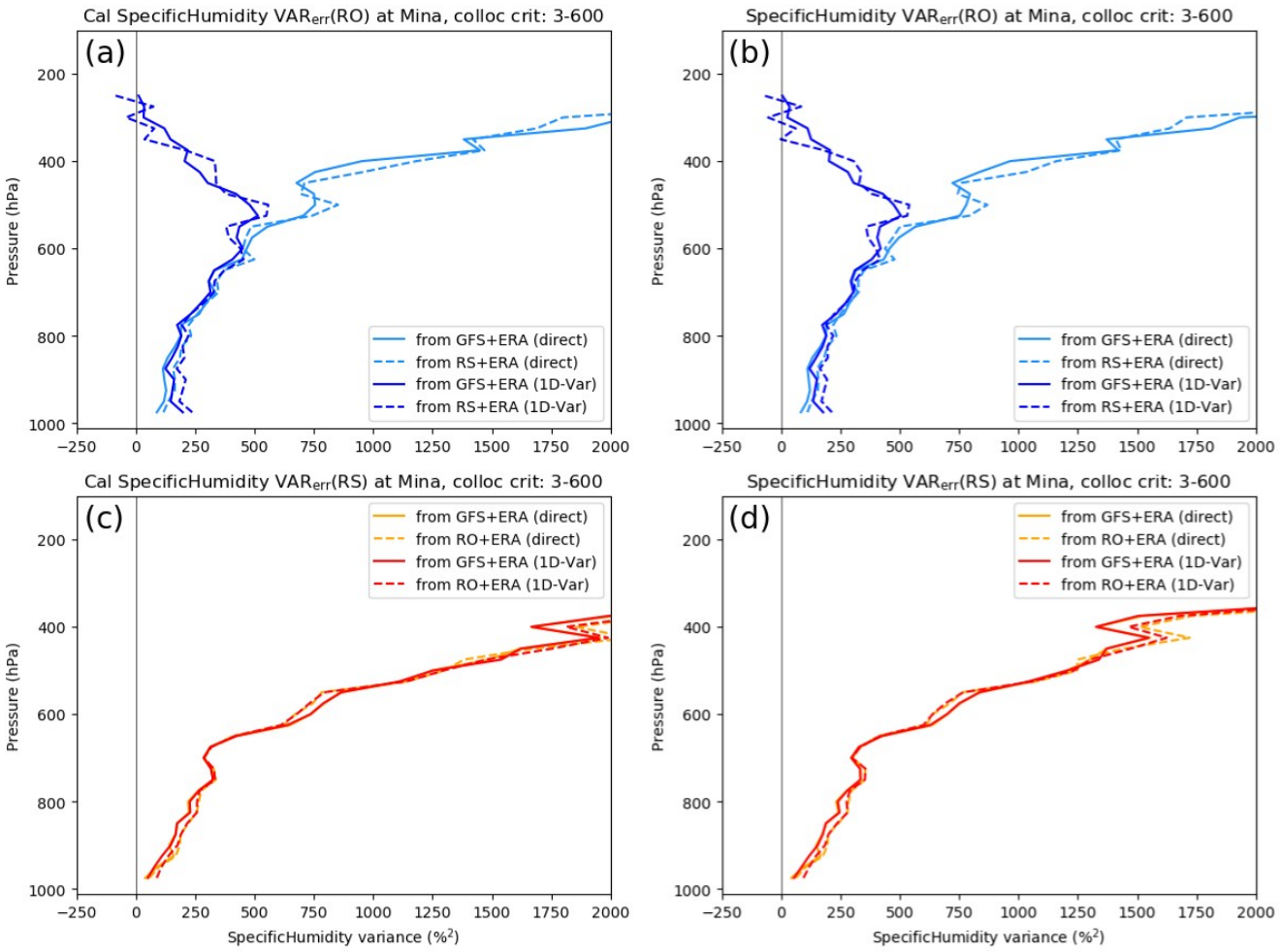

Fig. 1: Estimated RO and RS error variances for specific humidity at Minamidaitojima (Japan) using calibrated data as in the TC method (left) and the uncalibrated data as in the 3CH method (right). For the TC method, the RO, RS and GFS data sets are calibrated with respect to ERA as the reference data set. The following combinations of the 4 data sets are used: (ERA, RO, RS), (ERA, RO, GFS), and (ERA, GFS, RS).

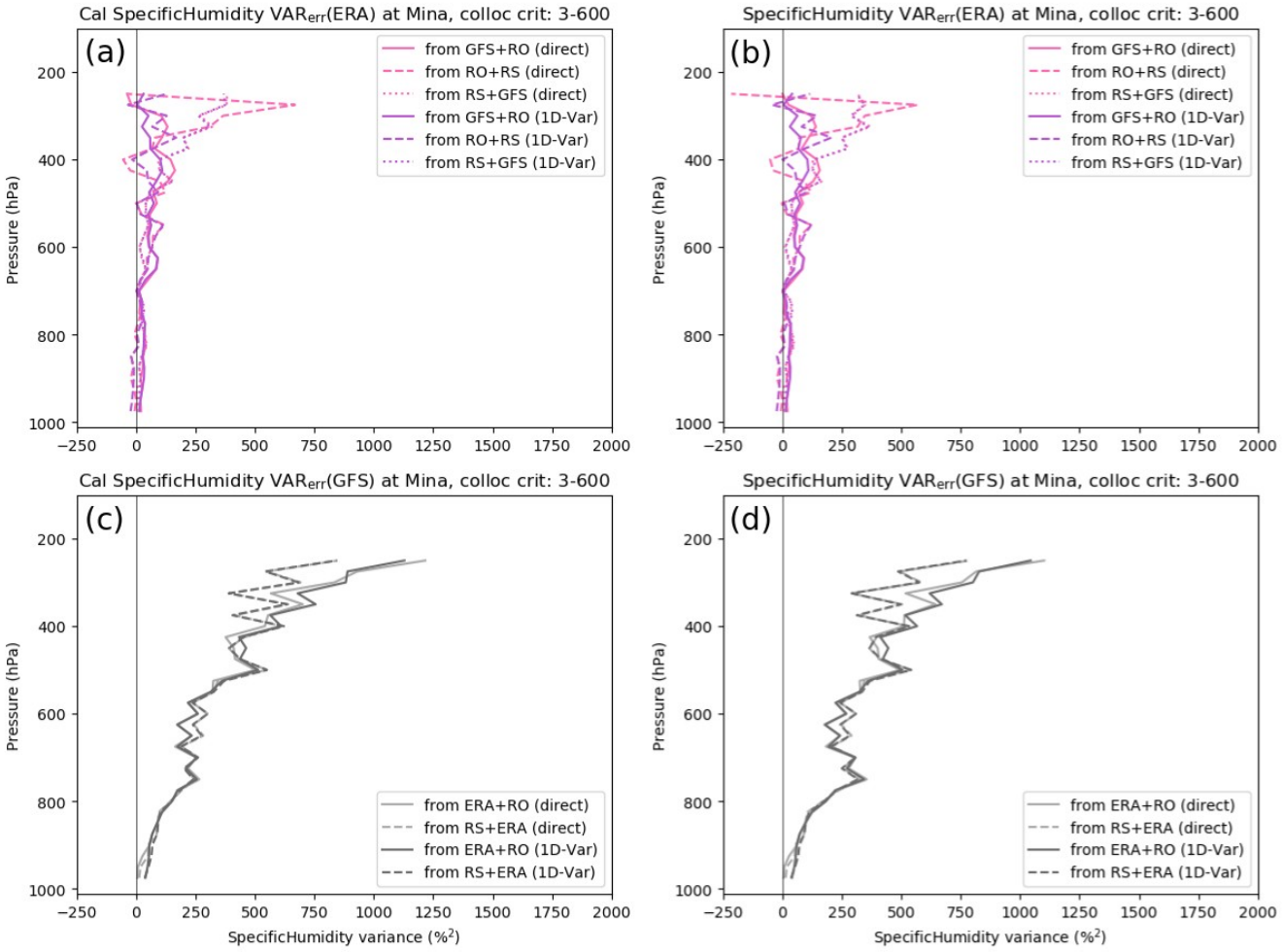

Fig. 2: Estimated ERA and GFS specific error variances for ERA at Minamidaitojima (Japan) using the triple co-location (TC) method (left) and the three-cornered hat (THC) method (right). For the TC method, the RO, RS and GFS data sets are calibrated with respect to ERA. The following combinations of the 4 data sets are used: (ERA, RO, RS), (ERA, RO, GFS), and (ERA, GFS, RS).

While 3CH and TC are similar in many respects, there are good reasons to use TC rather than 3CH when characterizing errors in geophysical data. Therefore, to address these concerns, I recommend the authors reframe their analysis using TC rather than 3CH. The differences between the techniques are relatively small, but significant, and warrant a substantial rewrite of the manuscript, in my view.

It would be interesting to do a detailed comparison of the 3CH and TC methods. However, given the small differences of the error variances computed of the TC and 3CH methods (as depicted in Figs. 1 and 2 above), there is no reason to redo our entire study using a different method. The 3CH method is well established for geophysical systems (O'Carroll et al., 2008; Valty et al., 2013) and the realistic results shown in our paper speak for themselves. As discussed in the previous response, our error model in paper AMT-2018-75 showed that biases have a relatively small effect in the 3CH method, a result that is supported by the similarity in the results comparing the 3CH and TC methods (Figs. 1 and 2 above)

We have addressed this issue, as discussed above, by adding a new section in Appendix A that compares the TC and 3CH methods using a subset of our data, and shows that the two methods give very similar results.

Specific comments Line 7, p 16: ". . .indicate that our estimates are reasonable and consistent with these studies." It seems unreasonable to be making this claim given the estimates in this study vary enormously. It is easy for an imprecise estimate to be consistent with previous studies, but this is not particularly informative.

To our knowledge, this is this first study of its kind comparing multiple error variance estimates by using several different combinations of data sets at the same locations and time using the 3CH method. As far as we know, there are no previous studies of any kind that estimate vertical profiles of the errors associated with specific humidity, relative humidity, temperature and refractivity estimates over the entire range of the troposphere as done in our paper. In addition, as discussed in the response to Reviewer 1, the error characteristics of the models, radiosondes and RO are likely to vary with latitude, longitude, seasons and year (as the models change, radiosondes use different sensors, and even RO data and processing change). Thus there are no standards for comparison. However, as we show, the results we obtain are reasonable and consistent with other studies of individual observing systems and the consistency in the three estimates for each observations and the clear differences in the error estimates of the different data sets support the validity of the results.

References

Stoffelen, Ad. "Toward the True Near-Surface Wind Speed: Error Modeling and Calibration Using Triple Collocation." Journal of Geophysical Research 103, no. C4 (1998): 7755–66.

Gruber, A., C. -H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner. "Recent Advances in (Soil Moisture) Triple Collocation Analysis." International Journal of Applied Earth Observation and Geoinformation, Advances in the Validation and Application of Remotely Sensed Soil Moisture - Part 1, 45, Part B (March 2016): 200–211. https://doi.org/10.1016/j.jag.2015.09.002.

Reviewer 3: Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-487, 2018 Received and published: 1 May 2018

General comments:
This is a very well written paper describing how one can estimate the error variances of different datasets of atmospheric profiles using the differences between three or more independent datasets of the same variable. The method ("N-cornered hat") is used to estimate the error variances in tropospheric profiles of four variables in five datasets at four locations in Japan and in Guam. The results indicate that the main assumption of neglecting error correlations between the datasets is reasonably valid. I enjoyed reading the paper and have only a few minor specific comments and technical corrections.

Thank you for your remarks and the detailed comments below.

Specific comments:

Page 3, line 4: "Paper 1" is only referred to once a few lines below (line 9). Thus it seems overkill to introduce it as "Paper 1" here. Perhaps in line 9 you can just say something like "Because the focus in Rieckh et al. (2017) was ...".

We agree and deleted the “Paper 1” reference.

Page 3, section 2.1. Please clarify if you are using analysis or forecast products from ERA-Interim. If you are using forecast products, that would be another good argument for small correlations, since the ERA forecasts only contain earlier observations via the assimilation, and are therefore independent of the observations that they are compared to.

We are using the analysis products from ERA-Interim and now state that on Page 3 Line 8 in the revised manuscript.

Figs. 6-9: The results without RO (those involving only RS,GFS,ERA) are listed twice in b,c,d panels. I would have expected identical curves, but there are small differences (e.g. refractivity from GFS,ERA,RS below 900 hPa in Fig 9c - light and dark gray dotted curves). Why is that?

This is a subtle point. The differences in the results using only RS, GFS, and ERA as pointed out above are due to slightly different sample sizes associated with the two RO retrievals in the mid and lower troposphere. There are two estimates of the error variances for RO because we estimate the error variance for RO using the Direct and 1D-VAR humidity retrievals of RO. This is discussed in the first paragraph of Section 5.1. For each RO retrieval, all four data sets (RS, RO, ERA, and GFS) are co-located. The data ensemble at each level is used only if data from all four sources are available on that level. The 1D-VAR RO data set extends to lower levels than the Direct RO data set because of the way the 1D-VAR is calculated. We now show this in a revised Figure 1 of the revised manuscript. Therefore, the number of data ensembles (including GFS, ERA and RS) available per level is slightly higher for all comparisons using the 1D-VAR retrieval compared to the number using Direct retrievals. These slightly different data sets give slightly different error variance estimates for the two (GFS,RS and ERA) ensembles.

Page 15, line 26: "... mean of the six estimates of the error variances ...". But it looks like the light and dark results are separated. Is the mean only over three estimates each? How is the standard deviation with only three estimates taken? For small samples, it becomes important that the denominator in the expression for the sample variance is more correctly written as N-1 (N=3 here). This would also alter the shaded areas in the appendix.

The reviewer is correct, this is the mean of the three estimates. We corrected the text on page 13 line 8 and the caption of Fig. 10.

Regarding the comment on how we computed the standard deviation, we used N, not N-1 in the original paper. We corrected this to use N-1 (two) in the revised paper (updated Figs. 10 and B1-B4). The results were not significantly different. We included the equation used to compute the STD in a footnote on page 13.

Page 21, line 30: "... dry and wet water vapor biases ...". Needs reformulation.

We reworded this to "...uses a radiosonde that is thought to have large water vapor biases...."

Page 24, line 26: Shouldn't it be Eqs. (3), (5), and (6)? And the same for the following eqs.

The reviewer is correct. However, we removed this entire section of the appendix (please see next comment).

However, eqs. (10), (11), and (12), with (13), (14), and (15) inserted are not additional independent equations (as you also write). For example, (10) with (13) inserted is equivalent to (7) + (8) - (9). In general, you could get many more (infinitely many) equations if you don't care that they are dependent, namely A(7)+B*(8)+C*(9), where A, B, and C are any numbers, except those where A+B+C=0. Thus, there are not only six different ways. There are three independent ways and infinitely many if you also count dependent ones that can be formed by linear combinations of the first three. I think you need to make clear that there is nothing unique about the three additional equations that you choose. As it is, one could get the impression that they are in some way special (in line 5 you write "the full six equations" and in line 18 "the remaining three estimates"). Perhaps they are special if one can show that they give rise to the smallest possible standard deviation of the variance estimates. I don't know if that is the case. In the first three independent equations, there are 3/2 MS terms, and 3 COV terms involved. In the additional three equations there are 5/2 MS terms and 5 COV terms. The more terms there are, the larger the standard deviation of the variance estimates will become, at least potentially. Interesting stuff!*

We agree with this comment. And indeed the standard deviations of the estimated error variances using the three linearly dependent equations are greater than those from the three linearly dependent equations, as stated in line 6 page 29 of the original manuscript. However, we decided to eliminate the results from the three linearly dependent equations in the revised paper. We added the following paragraph in Appendix A:

"As noted by an anonymous reviewer, it is possible to derive infinitely many linearly dependent equations by combining Eqs. (A8)-(A10) in different ways using the form M_1 x Eq. (A8) + M_2 x Eq. (A9) + M_3 x Eq. (A10) where M_1 , M_2 and M_3 are any numbers except those for which $M_1 + M_2 + M_3 = 0$. We did not pursue this possibility in this paper, but instead used the three linearly independent equations only in our estimates of error variances."

Page 25, line 15-16: I do not understand this part of the sentence: "...the set of observations in the pairs (RO,ERA), (RO,GFS), (GFS,ERA), (RO,RS), (RS,ERA) and (RS,GFS) are the same (they are in our case)..." What are the set of observations here?

We clarified this by writing: "If all the neglected covariance terms were in fact identically zero and the sample size was very large (much larger than our sample size), all three estimates of the error variances would be the same."

Technical corrections:
Page 2, line 3: model -> modeled

Done

Page 3, line 14: Missing "a" in front of global.

Done

Page 11, line 6: Acronym STD should be written out first time.

Done

Page 12, line 32: that -> than

Done

Page 15-16, lines 30,1-2: Should RS and RO be switched in the text here? RO is the one that oscillates between 0.1 and 0.3 (% squared). RS is fairly constant about 0.1.

Yes. We corrected and clarified the description in the revised text.

Page 21, line 17: Missing "to" in front of RO.

Corrected

Page 22, line 27: Vogelznang -> Vogelzang

Corrected

Page 25, line 18: "in are in"?

A typo-we deleted the first “in”.

References used in our responses:

Anlauf, H., Pingel, D., and Rhodin, A.: Assimilation of GPS radio occultation data at DWD, Atmos. Meas. Tech., 4, 1105–1113, https://doi.org/10.5194/amt-4-1105-2011, 2011.

Blackmore, T., O,Carroll, A., Saunders, R., and Aumann, H.H.: A comparison of sea surface temperature from the AATSR and AIRS instruments. Met. Office Forecasting Research Technical Report No. 495, 2007. (Available from nwp-publications@metoffice.gov.uk)

Burrows, C., Healy, S., and Culverwell, I.: Improving the bias characteristics of the ROPP refractivity

and bending angle operators, Atmos. Meas. Tech., 7, 3445–3458, https://doi.org/10.5194/amt-7-3445-2014, 2014.

Cucurull, L., Derber, J., and Purser, R.: A bending angle forward operator for global positioning system radio occultation measurements, J. Geophys. Res., 118, 14–28, https://doi.org/10.1029/2012JD017782, 2013.

Gilpin, S., Rieckh, T, and Anthes, R.: Reducing representativeness and sampling errors in radio occultation-radiosonde comparisons., Atmos. Meas. Tech., 11, 1-16, https://doi.org/10.5194/amt-11-1-2018, 2018a.

Gilpin, S., Anthes, R., and Sokolovskiy, S: Sensitivity of forward-modeled bending angles to vertical interpolation of refractivity for radio occultation data assimilation. To be submitted to Atmos. Meas. Tech., 2018b.

Healy, S. and Thépaut, J.-N.: Assimilation experiments with CHAMP GPS radio occultation measurements, Mon. Wea. Rev., 132, 605–623, 2006.

O'Carroll, A.G., Eyre, J.R., and Saunders, R.S: Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. J. Atmos. Oceanic Tech., 25,1197-1207, https://doi.org/10.1175/2007JTECHO542.1, 2008.

Valty, P., de Viron, O., Panet, I., Camp, M. V., and Legrand, J.: Assessing the precision in loading estimates by geodetic techniques in Southern Europe, Geophys. J. Int., 194, 1441–1454, https://doi.org/https://doi.org/10.1093/gji/ggt173, 2013.

Estimating observation and model error variances using multiple data sets

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

[1]COSMIC Program Office, University Corporation for Atmospheric Research, Colorado, U.S.A.
[2]Wegener Center for Climate and Global Change, University of Graz, Graz, Austria

Correspondence: Richard Anthes (anthes@ucar.edu)

Abstract. In this paper we show how multiple data sets, including observations and models, can be combined using the "three cornered hat~~method" ~~" (3CH) method to estimate vertical profiles of the errors of each system. Using data from 2007, we estimate the error variances of radio occultation, radiosondes, ERA-Interim, and GFS model data sets at four radiosonde locations in the tropics and subtropics. A key assumption is the neglect of error covariances among the different data sets, and we examine the consequences of this assumption on the resulting error estimates. Our results show that different combinations of the four data sets yield similar relative and specific humidity, temperature, and refractivity error variance profiles at the four stations, and these estimates are consistent with previous estimates where available. These results thus indicate that the correlations of the errors among all data sets are small and the 3CH method yields realistic error variance profiles. The estimated error variances of the ERA-Interim data set are smallest, a reasonable result considering the excellent model and data assimilation system and assimilation of high-quality observations. For the four locations studied, RO has smaller error variances than radiosondes, in agreement with previous studies.

1 Introduction

Estimating the error characteristics of any observational system or model is important for many reasons. Not only are these errors of scientific interest, they are important for data assimilation systems and numerical weather prediction. In many modern data assimilation schemes, observations of a given type are weighted proportionally to the inverse of their error variance (e.g. Desroziers and Ivanov, 2001).

Kuo et al. (2004) and Chen et al. (2011) used the difference between radio occultation (RO) observations ~~of a variable X (e.g.refractivity)~~ and short-range model forecasts ~~of X~~ to estimate the error of the RO observations, using the concept of apparent or perceived errors, defined by

$$X_{AE} = X_{RO} - X_{fcst} \quad (1)$$

where X_{AE} is the apparent error of the RO observation and X_{RO} and X_{fcst} are the RO observations and model forecast values, respectively.

The error variance σ_a^2 of the apparent error is given by

$$\sigma_a^2 = \frac{1}{n} \sum \frac{X_{AE}^2}{N} X_{AE}^2 \tag{2}$$

where ~~N~~ n is the number of samples of observed and ~~model~~ modeled RO at the same location and time.

The relationship between the apparent error variance σ_a^2, the observational error variance σ_o^2, and the forecast error variance σ_f^2 is given by:

$$\sigma_a^2 = \sigma_o^2 + \sigma_f^2 - 2\mathrm{COV}_{\mathrm{err}}(X_{\mathrm{RO}}, X_{\mathrm{fcst}}) \tag{3}$$

where the $\mathrm{COV}_{\mathrm{err}}$ term is the error covariance between the observations and the forecasts. If the error variance of the forecast σ_f^2 is estimated independently, the observational error variance can be estimated from the apparent error variance, under the assumption that the observational errors are uncorrelated with the forecast errors (in which case the $\mathrm{COV}_{\mathrm{err}}$ term in Eq. (3) is zero).

$$\sigma_a^2 = \sigma_o^2 + \sigma_f^2 \tag{4}$$

We note that the apparent errors are the same as the (O – B) (observation minus background) or innovations as used in data assimilation methods and studies (Chen et al., 2011).

As discussed by Kuo et al. (2004) and Chen et al. (2011), the forecast error variance can be estimated by two alternative methods, the ~~"NMC method "~~ NMC method (Parrish and Derber, 1992) or the Hollingsworth and Lönnberg (1986) method. Kuo et al. (2004) used both methods to estimate the observational errors of RO refractivity using the NCEP AVN model. Chen et al. (2011) used the NMC method and Weather and Research Forecast Model (WRF) to estimate the forecast error variance and then the RO refractivity error variance.

In this paper, we ~~develop a method for estimating~~ estimate the error variances of multiple data sets using the "three-cornered hat" (3CH) method (Gray and Allan, 1974) . Unlike the apparent error method, this method does not require independent estimates of the error variance of a forecast; it uses the differences between combinations of three data sets~~using the "three cornered hat" method (description~~ . The 3CH method is described in Appendix A ~~)~~along with the closely related "triple co-location method" (Stoffelen, 1998) . The data sets may be either different model or observational data and estimates of the error variances of all the data sets are computed by the method. We compare three observational data sets (two versions of ~~radio occultation~~ RO retrievals and radiosondes (RS)) and two model data sets at four locations in the tropics and subtropics to estimate the error variances of all five data sets. We find that the results are consistent with each other and with previous error estimates, where available.

2 Discussion of data sets

We use five data sets from an entire year (2007) in this study. ~~Rieckh et al. (2017, hereafter Paper 1) ,~~ Rieckh et al. (2018) extensively studied the properties of these data sets and their daily variability over 2007 in the tropical and sub-tropical western Pacific. They are described in more detail there, but are summarized briefly here for convenience.

We chose 2007 for the year of our study because the number of COSMIC (Constellation Observing System for Meteorology, Ionosphere and Climate) RO observations was near a maximum at this time. Because ~~our~~ the primary interest in ~~Paper 1~~ Rieckh et al. (2018) was the evaluation of water vapor observations and model analyses in challenging tropical and subtropical environments, we chose one ~~radiosonde (RS)~~ RS station in the deep tropics and three Japanese stations in the subtropics. Because of our focus on water vapor, we carry out the analysis from 1000 to 200 hPa.

2.1 ERA-Interim

The ERA-Interim (hereafter ERA) reanalysis is a global model reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Dee et al., 2011). Information about the current status of ERA-Interim production, availability of data online, and near-real-time updates of various climate indicators derived from ERA-Interim data ~~,~~ can be found at ~~https://www.ecmwf.int/en/research/climate-reanalysis/reanalysis-datasets/era-interim~~https://www.ecmwf.int/en/research/climate-reanalysis/reanalysis-datasets/era-interim.

~~The ERA~~ We use the ERA analysis product, which assimilates both RS and RO data for the entire year of 2007; hence some correlation of model, RS, and RO errors is likely. However, there are many other observations going into the ERA reanalysis and model correlations with any one observational data set are likely to be small.

2.1.1 ~~NCEP Global Forecast System (GFS)~~

2.2 NCEP Global Forecast System (GFS)

The Global Forecast System (GFS) is a forecast model produced by the National Centers for Environmental Prediction (NCEP). Data are available for download through the NOAA National Operational Model Archive and Distribution System (NOMADS). Forecast products and more information on GFS are available at ~~https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-~~

~~Prior to January 2003, the GFS was known as the Aviation model (AVN), which was one of the models used by Kuo et al. (2004) in their estimation of RO errors using the apparent error method~~https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs.

The GFS assimilated RS observations for the entire year 2007, but began assimilating RO data on 1 May, 2007, along with many other changes to the model and analysis system (Cucurull and Derber, 2008; Kleist et al., 2009). Thus the GFS and RS and RO errors are also likely correlated to some degree. However, we computed vertical profiles of the correlation coefficients for RO and GFS refractivity, temperature, specific humidity, and relative humidity in the two months before and after May 1, 2007 when the GFS started assimilating RO data and found little differences, so the error correlations between RO and GFS are likely small.

2.2.1 ~~Radio occultation observations~~

2.3 Radio occultation observations

The RO observations used in this study are re-processed data obtained from the UCAR COSMIC Data Analysis and Archive Center (CDAAC). Two methods for estimating the temperature and water vapor from the RO refractivity are used. In the direct method, the GFS temperature is used in the Smith and Weintraub (1953) equation

$$N = 77.6\frac{P}{T}\frac{p}{T} + 3.73 \cdot 10^5 \frac{e}{T^2} \quad (5)$$

to compute water vapor pressure ~~e~~ e from the observed ~~N~~ refractivity N and GFS temperature ~~T~~T.

A one-dimensional variational (1D-VAR) method is also used to estimate ~~temperature and water vapor pressure from N~~T and e from N. The 1D-VAR method uses an a-priori state of the atmosphere (background profile) and an observed RO ~~refractivity~~ N profile to minimize a quadratic cost function. At CDAAC, an ERA-Interim profile is used as background, which is interpolated to the time and location of the RO observation (accounting for tangent point drift during the occultation). The humidity retrieval allows an error for both ~~T and e~~T and e, but only a very small error for bending angle/refractivity. Specific humidity ~~q~~ q is then computed from the derived ~~water vapor pressure e~~e.

2.3.1 ~~Radiosonde observations~~

2.4 Radiosonde observations

~~Radiosonde~~ RS data from Guam and three Japanese stations are used in this comparison. The ~~radiosondes~~ RS data are given on nine main pressure levels between 1000 hPa and 200 hPa, plus additional levels if atmospheric conditions are variable. The four stations use the following sensors: Guam: VIZ/Sippican B2; Ishigakijima: Meisei; Minamidaitojima: Vaisala RS92; and Naze: Meisei. They are launched twice daily in the hour before noon and midnight, UTC.

Guam is located in the deep tropics at 13.7°N 144.8°E. Ishigahijima (hereafter called Ishi), Minamidaitojima (hereafter called Mina), and Naze are located relatively close together in the western Pacific subtropics south of Japan and northeast of Taiwan:

Naze: Naze/Funchatoge (Kagoshima) 28.4°N 129.4°W

Mina: Minami-daitojima (Okinawa) 25.6°N 131.5°W

Ishi: Ishigakijima (Okinawa) 24.2°N 124.5°W

~~The RO observations used in the comparisonsare those~~

2.5 Co-location of the data sets

The locations of the four radiosonde stations are chosen for the comparisons. We use RO observations that are located within 600 km and 3 hours of the radiosonde ~~launch. The model observations closest to the four radiosonde stations respectively are~~

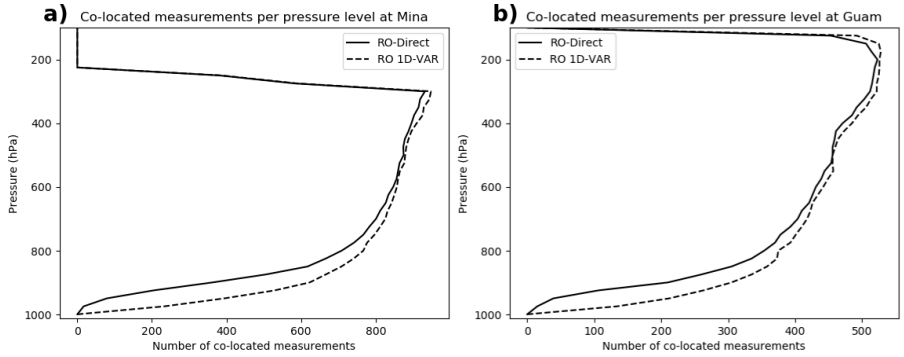

Figure 1. Number of ~~RO refractivity profiles matching the co-location criteria~~ co-located measurements for ~~Mina~~ (a) Mina and ~~Guam~~ (b) ~~during the year 2007.~~ Guam. These are also the ~~number of samples~~ sample numbers in the calculations of the estimated error variances for the ~~five~~ data sets~~(RO Direct and~~ . When using the RO 1D-VAR, ~~RS, ERA and GFS)~~the number of co-locations is slightly higher than for the RO-Direct throughout the profile due to the way the 1D-VAR is computed.

~~used. Corrections for the spatial separation are made using the model data . The details are described by Gilpin et al. (2017, in preparation)~~ ~~l~~ CDAAC provides GFS and ERA profiles that are already linearly interpolated in space and time to the RO location and time. These interpolated profiles, along with the RO observations, were corrected for their time and spatial differences from the radiosonde data using a model correction algorithm (described in Gilpin et al., 2018) . Thus the effect of spatial and temporal differences among the data sets is expected to be minor.

The refractivity for the radiosonde and model data is computed from Eq. (5) using the pressure, temperature, and water vapor from these data. Normalized differences are computed for all combinations of the data sets (RO–ERA, RO–GFS, GFS–ERA, RS–ERA, RS–GFS, RS–RO), where RO is either the RO-Direct or the RO 1D-VAR data. The ERA annual mean for 2007 at each RS station is used to normalize the differences in the data sets associated with that station. We consider the differences among ~~the five~~ all data sets for four variables: refractivity (~~N~~N), temperature (~~T~~T), specific humidity (~~q~~)q), and relative humidity (RH).

2.5.1 ~~Number of samples~~

2.6 Number of samples

The number of samples is limited by the number of RO observations that are within the co-location criteria of ~~three~~ 3 hours and 600 km. Figure 1 shows the number of ~~RO profiles~~ data samples per pressure level that meet these criteria during 2007 at Mina (the numbers at Ishi and Naze are similar) and Guam. The ~~number of samples at the Japanese stations is a maximum of approximately 900 at 300hPa. It decreases rapidly above 300hPa to zero at 200hPa because of the~~ sharp cut-off on the top of the profiles is due to limited RS data availability at high altitudes. The smooth transition to lower numbers at the bottom results from a decrease in the number of ~~RS observations above~~ RO observations with lower altitudes in the mid and lower

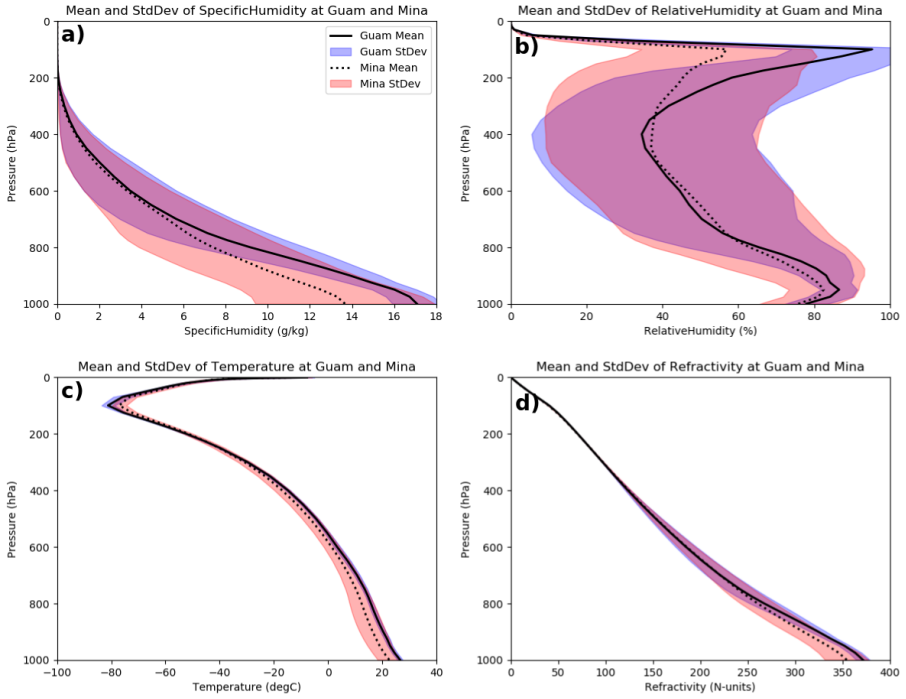

Figure 2. The mean ERA profiles over 2007 at Guam and Mina of specific humidity ~~q~~ q (a), relative humidity RH (b), temperature ~~T~~ T (c), and refractivity ~~N~~ N (d). The standard deviations about the mean profiles are indicated by the shading.

troposphere. The number of samples at the Japanese stations is a maximum of approximately 900 at 300 hPa. The number decreases to about 100 at 950 hPa~~at the three Japanese stations~~. At Guam, the number ranges from a maximum of about 500 at 200 hPa to about 50 at 950 hPa~~at Guam~~. Thus the effect of the limited sample size will be greatest for the Japanese stations above 300 hPa and for all four stations below 900 hPa where the sample size is less than 500.

2.6.1 ~~Mean ERA profiles for 2007 and example of profiles and normalized difference profiles~~

2.7 Mean ERA profiles for 2007 and example of profiles and normalized difference profiles

Before showing the statistical comparisons of the normalized differences between the data sets and their estimated errors, we present the mean ERA profiles of ~~q~~ q, RH, ~~T and N~~ T and N at Mina and Guam for the year 2007 (~~Figure~~Fig. 2). The standard deviations are shown by the shading around each mean profile. As shown by ~~Figure~~Fig. 2, the water vapor (especially relative humidity) shows the greatest variability over the year. The variability in specific humidity, temperature, and refractivity is greater at Mina, which is located in the subtropics, than Guam, which is located in the deep tropics.

We next present a single example of soundings from the five data sets, to illustrate how the profiles of the normalized differences of the variables (which we use in all the following calculations) compare to the actual profiles. Figure 3 illustrates the ~~q~~ q,

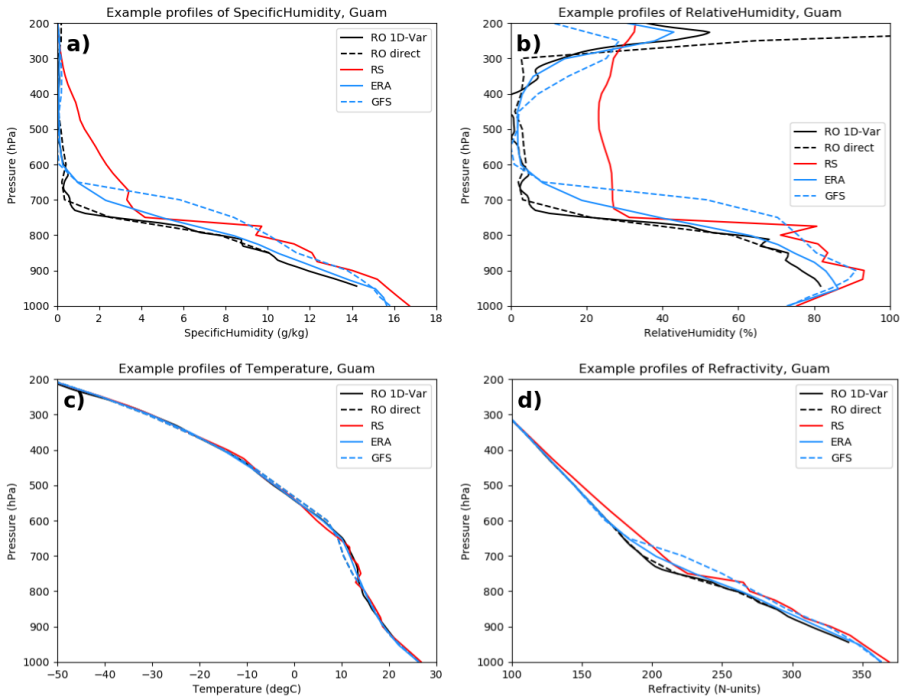

Figure 3. Profiles of specific humidity (a), relative humidity (b), temperature (c) and refractivity (d) for the five data sets for 13 January, 2007 at 00:23 UTC.

RH, ~~T and N~~ T, and N profiles from 13 January 2007 at approximately 00 GMT and ~~Figure~~Fig. 4 illustrates the corresponding profiles of the normalized differences of the variables from ERA, for example $(q - q_{\mathrm{ERA}})$/~~CLIMO, where CLIMO~~ $\bar{q}$, where $\bar{q}$ is the 2007 mean ERA value of ~~the variable~~q.

A comparison of ~~Figures~~Figs. 3 and 4 shows that the normalized difference profiles highlight the similarities and differences of the five data sets better than the actual profiles, especially in the upper troposphere. The magnitudes of the normalized differences are the same order of magnitude at all levels, whereas the differences in the actual profiles can vary by more than an order of magnitude from the lower to the upper troposphere. Figure 4 shows that typical percentage differences between data sets are ~50 % for ~~q~~ q and RH, 0.5 % for ~~T~~T, and 5 % for ~~N~~N.

3 Derivation of error variances

In this section we summarize the derivation of the equations relating the error variances and covariances among the ~~five~~ data sets. The complete derivation and a discussion of the limitations is given in Appendix A.

The error variance of a variable X (e.g. ~~q~~q, RH, ~~T or N~~T or N) is defined as

$$\mathrm{VAR_{err}(X)} = \frac{1}{n}\sum(\mathrm{X} - \mathrm{X_t\,True})^2 \,/\mathrm{N} = \frac{1}{n}\sum \mathrm{X_{err}^2} \tag{6}$$

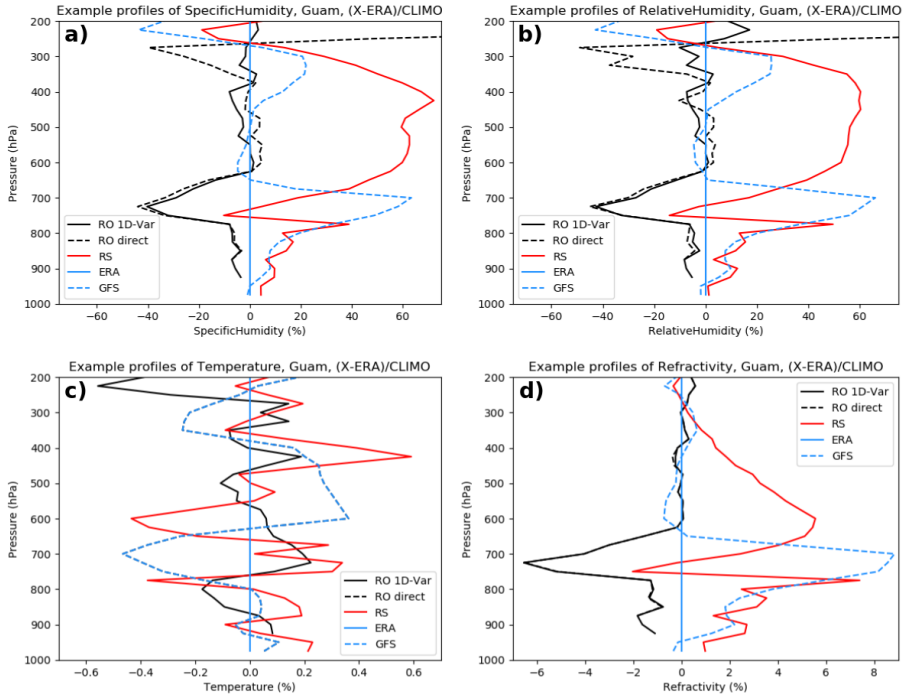

Figure 4. Same as ~~Figure~~Fig. 3 except for normalized differences from ERA.

where ~~X_t~~ True is the true (but unknown) value of X and the summation is over ~~N~~ n samples.

As shown in Appendix A, we can derive three different linearly independent equations for estimating the error variance of any data set, assuming that the error covariances among all the data sets are negligible compared to the differences in the observed mean square (MS) differences between the data sets. ~~Three more equations can be derived by using linear combinations of the first three. Appendix A provides examples of the error estimates using the six and three equations. For example, the~~ The three complete (and exact) linearly independent solutions for estimating the error variance of RO are

$$2\text{~~VAR(RO – True) =~~ VAR}_{\text{err}}\text{(RO)} = \text{~~MS(RO – ERA) + MS(RO – GFS) - MS(GFS – ERA)~~MS(RO – ERA) + MS(RO – GFS) – MS(GFS –}$$
$$+ \text{~~2 COV(RO, ERA) + COV(RO, GFS) – COV(GFS, ERA)~~}2\left[\text{COV}_{\text{err}}\text{(RO, ERA)}+\text{COV}_{\text{err}}\text{(RO, GFS} \right. \tag{7}$$

$$2\text{~~VAR(RO – True) =~~ VAR}_{\text{err}}\text{(RO)} = \text{~~MS(RO – ERA) + MS(RO – RS) - MS(RS – ERA)~~MS(RO – ERA) + MS(RO – RS) – MS(RS – ERA}$$
$$+ \text{~~2 COV(RO, ERA) + COV(RO, RS) – COV(RS, ERA)~~}2\left[\text{COV}_{\text{err}}\text{(RO, ERA)}+\text{COV}_{\text{err}}\text{(RO, RS)}-\text{C} \right. \tag{8}$$

$$2\text{VAR(RO – True)} = \text{VAR}_{\text{err}}(\text{RO}) = \text{MS(RO – GFS)} + \text{MS(RO – RS)} - \text{MS(RS – GFS)} \text{MS(RO – GFS)} + \text{MS(RO – RS)} - \text{MS(RS – GFS)}]$$

$$+ 2\,\text{COV(RO, GFS)} + \text{COV(RO, RS)} - \text{COV(RS, GFS)} 2\left[\text{COV}_{\text{err}}(\text{RO, GFS}) + \text{COV}_{\text{err}}(\text{RO, RS}) - \text{C}\right. \quad (9)$$

where RO (or ERA, GFS, RS) corresponds to the value of X as estimated by RO (or ERA, GFS, RS) , True corresponds to the true (but unknown) value of X and MS denotes the mean square difference between the values from two data sets (e.g. RO – ERA).

We use Eq. (7)–(9) to provide three independent estimates of VAR(RO – True$_{\text{err}}$(RO) by neglecting the COV$_{\text{err}}$ terms in each equation. The assumption that the error covariances are small compared to the difference in variances between the data sets is similar to the assumption used in the apparent error method that the errors of the observations and model forecasts are uncorrelated. Of course in In general the COV$_{\text{err}}$ terms are not zero; thus we will examine the validity of this assumption by checking whether the various estimates of the error variances from the three equations are consistent with each other and reasonable compared to other independent studies that estimate error variances in other ways. In a related paper (Rieckh and Anthes, 2018) we examine the effect of various degrees of error correlations between two of the three data sets using an error model.

The same procedure can be used to derive three equations for estimating the error variances for the other three data sets, RS, ERA, and GFS (equations not shown here).

So for each of the five data sets, RO Direct and RO 1D-VAR, RS, ERA, and GFS, there are three independent ways to estimate their respective error variances. This is the "three cornered hat" method described in Appendix A. We note that it is possible that the estimated error variances from any of the three equations are negative because of the neglect of the COV$_{\text{err}}$ terms and the small sample size, especially above 300 hPa for the Japanese stations and below 800 hPa for all four stations (FigureFig. 1).

4 Comparison with previous studies for RO refractivity

We first compute the estimated error variance for RO refractivity using GFS and ERA data for comparison with the Kuo et al. (2004) and Chen et al. (2011) estimates of RO error variance to illustrate the three cornered hat 3CH method. In an analogy to the apparent error Eq. (4), with RO being the observation and ERA being the forecast

$$\text{MS(RO – ERA)} = \text{VAR(RO – True)} + \text{VAR(ERA – True)} \text{MS(RO – ERA)} = \text{VAR}_{\text{err}}(\text{RO}) + \text{VAR}_{\text{err}}(\text{ERA}) \quad (10)$$

which is Eq. (A2) in Appendix A with neglect of the COVterms.

$_{\text{err}}$ terms. We compute MS(RO – ERA) from the RO and ERA data sets (analogous to the apparent error variance σ_a^2 in Eq. (4)) and plot its square root as the black line in FigureFig. 5. Then we estimate VAR(RO – True$_{\text{err}}$(RO) using Eq. (7) and the data sets (RO – GFS) and (GFS – ERA), along with the apparent error MS(RO – ERA), neglecting the COV$_{\text{err}}$ terms.

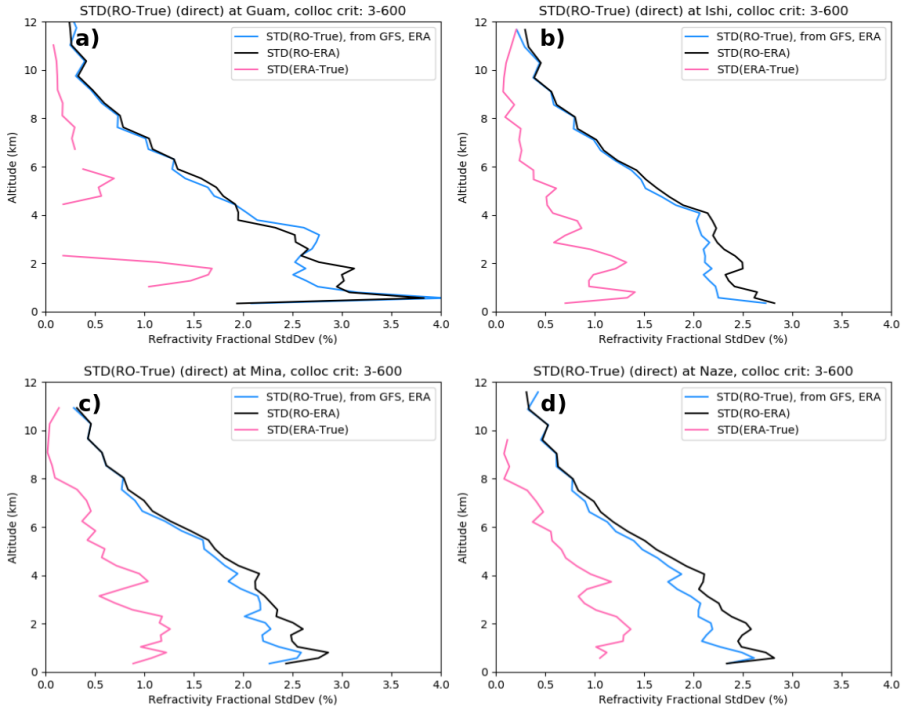

Figure 5. Standard deviations of the apparent error STD(RO – ERA) (black line), estimated RO error STD(RO – True) computed from Eq. (~~11~~7) (blue line) and ERA error STD(ERA – True) (pink line) for refractivity at a) Guam, b) Ishi, c) Mina, and d) Naze.

The square root of VAR~~(RO-True~~$_{err}$(RO) gives the ~~blue curve labelled STD(RO – True) in Figure 5~~standard deviation (STD) (Fig. 5, blue curve). Finally, the ERA error variance (analogous to the forecast error) is obtained by subtracting VAR~~(RO – True~~$_{err}$(RO) from MS(RO – ERA) using Eq. (10) above (pink line in ~~Figure~~Fig. 5). The gap in the computed ERA error STD in Fig. 5a occurs due to negative estimated error variance values, which can result from having a limited sample size, neglecting error covariance terms during computation, and having an error variance that is already close to zero (as is the case for ERA).

The results shown in ~~Figure~~Fig. 5 are quite similar to those from ~~Kuo et al. (2004, Figure 13) and Chen et al. (2011, Figure 3d)~~ Kuo et al. Chen et al. (2011, Fig. 3d) who used different models and different data sets. The STD of normalized RO refractivity errors are a maximum of between 2.0 and 2.5 % near the surface, decreasing to about 0.5 % at 10 km. ~~The shape of the profiles between 0 and 2km is also similar in the two methods, with a small local minimum in the profile at about 1km.~~ These similarities give credibility to both methods.

5 Calculation of the error variance terms using the multiple data sets

This section shows the estimated error variances for ~~N, q, T~~ N, q, T, and RH at one of the four stations (Mina) for the five data sets and summarizes the results for the other three stations (Naze, Ishi and Guam).

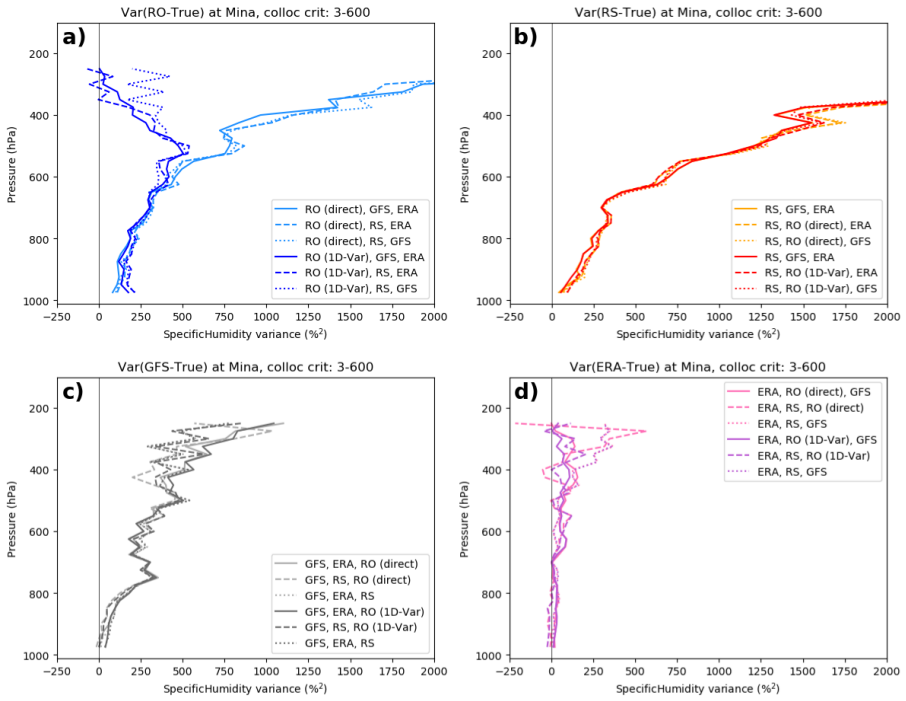

Figure 6. Estimated error variances (% squared) of specific humidity at Mina: a) RO, b) RS, c) GFS, and d) ERA.

5.1 Results for Mina

The following plots show the estimated error variances computed from Eq. (7), (8), or (9). Two RO data sets (Direct and 1D-VAR) are considered one at a time using the other three data sets. Thus we have two sets of error estimates for each data set: one using the ~~Direct RO~~ RO-Direct with RS, ERA, and GFS, and one using the ~~1D-VAR RO~~RO 1D-VAR with RS, ERA, and GFS. In the following plots, ~~results for six error estimates are indicated by the color. Darker~~ darker colors correspond to the three results using the ~~1D-VAR RO,~~ RO 1D-VAR and lighter colors correspond to the three results using the ~~Direct RO~~RO-Direct.

Figure 6 shows the results for specific humidity. Error variances are shown rather than STD because they are easier to interpret using the three equations used to derive them and because the STD are undefined for the occasional negative estimated error variance. Figure 6a shows the q error variance profiles for the two RO data sets (Direct and 1D-VAR). The direct method (use of GFS temperature in Eq. (5)) shows a steady increase of error variance with height, from about 100 $\%^2$ (STD ~10 %) at 950 hPa to 800 $\%^2$ (STD ~28 %) at 500 hPa and 2000 $\%^2$ (STD ~45 %) at 300 hPa. This is expected since the refractivity contains little information on water vapor above about 400 hPa and we are using an independent estimate of temperature, with no constraints on the water vapor retrieval. The q error variance profile for RO using the 1D-VAR method is similar to that of the direct method below 500 hPa, but reaches a maximum at about 500 hPa of about 500 $\%^2$ (STD ~22 %) and then decreases toward zero at 200 hPa. The 1D-VAR method uses the ERA-Interim fields as background and thus constrains the water vapor

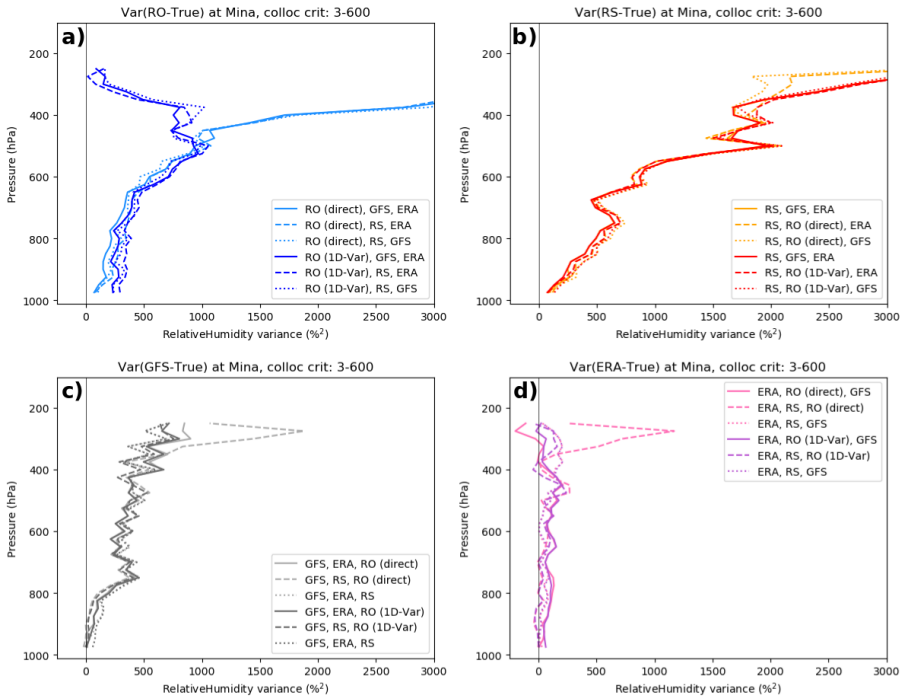

Figure 7. Estimated error variances (% squared) of relative humidity at Mina: a) RO, b) RS, c) GFS, and d) ERA.

profile retrieval at high altitudes. It is notable that the three equations used to estimate the error variance profiles agree closely and the difference among the three estimates is much smaller than the differences in the mean profiles using the two RO retrieval methods.

The RS specific humidity error variance profiles ~~associated with the radiosonde~~ at Mina (~~Figure~~Fig. 6b) show a similar behavior as the ~~direct RO~~RO-Direct, with a steady increase with height, exceeding ~~a VAR of~~ 2000 $\%^2$ (STD of $\sim$45 %) at 400 hPa. The STD of the RS are slightly larger than the two RO estimates below 600 hPa. The error variance estimates using the ~~Direct RO~~ RO-Direct (orange) and ~~1D-VAR RO~~RO 1D-VAR (red) are similar.

The error variance profiles from the two model sets (~~Figure~~Figs. 6c, d) are quite different. The GFS error variance is less ~~that the RO Direct and radiosonde~~ than the RO-Direct and RS error variances at all levels, and also less than the RO 1D-VAR error variance except above 300 hPa. Although there is more scatter, especially in the upper troposphere, the ERA profiles are different from all the other data sets in that they show only a small increase of error variance with height, from ~~a variance~~ near zero at the surface to up to a mean of about 100 $\%^2$ (STD $\sim$10 %) at 200 hPa. ~~The ERA profiles contain examples of the estimated error variances becoming negative. This is because the true values of the error variances are close to zero and so neglect of the error covariance terms can produce small negative values.~~

Figure 7 shows the estimated error variances of relative humidity. As with specific humidity, there is consistency among the estimates for the different data sets. The general behavior of the RH error variance profiles is similar to ~~those~~ that for q, as might

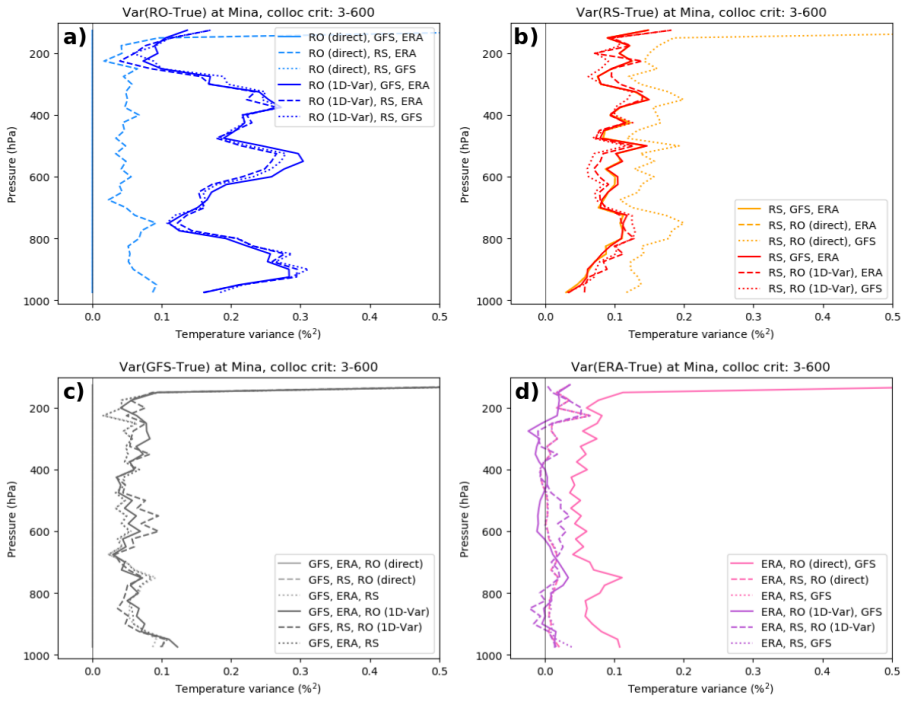

Figure 8. Estimated error variances (% squared) of temperature at Mina: a) RO, b) RS, c) GFS, and d) ERA.

be expected because the percentage variability of water vapor is greater than that of temperature at this subtropical location. Again, the estimated error variances of the RO derived RH are less than those of the ~~radiosondes~~ RS in the lower troposphere. The GFS error variances are smaller than the RO and RS variances, except for the RO 1D-VAR profile above 300 hPa, which is constrained by the ERA observations in the upper troposphere. The ERA error variances are significantly smaller than the other data sets, averaging between 50 and 200 $\%^2$ (STD ~~7–14~~7–14 %) throughout the troposphere.

Figure 8 shows the estimated error variances of temperature. Because the RO-Direct retrieval uses the exact GFS temperature, the results for the direct retrieval (light blue) using the (RO, GFS, ERA) and (RO, RS and GFS) are not meaningful in ~~Figure~~Fig. 8a (they are identically zero). The result from Eq. (8) (RO, ERA and RS), given by the dashed light blue line in ~~Figure~~Fig. 8a is valid, but in reality, this is an estimate of the GFS T error variance, and it is in fact very similar to the profiles in ~~Figure~~Fig. 8c.

The RO 1D-VAR results for temperature from all three equations give somewhat larger results (~~Figure 9a~~Fig. 8, dark blue profiles). The estimated error variance profiles oscillate between 0.1 and 0.3 $\%^2$ (STD 0.3 to 0.55 %). For a temperature of 300 K, these correspond to 0.9 to 1.65 K.

The ~~radiosonde~~ RS temperature error variances (~~Figure~~Fig. 8b) vary between 0.05 and 0.15 $\%^2$ (STD 0.2 to 0.4 % or 0.6 to 1.2 K for T=300 K). The GFS temperature error variances are a little lower, averaging around 0.05 ~~–0.10~~ to 0.10 $\%^2$ (STD 0.2 ~~-0.3~~to 0.3 %), while the ERA estimated temperature error variances average close to zero (~~Figure~~Fig. 8d).

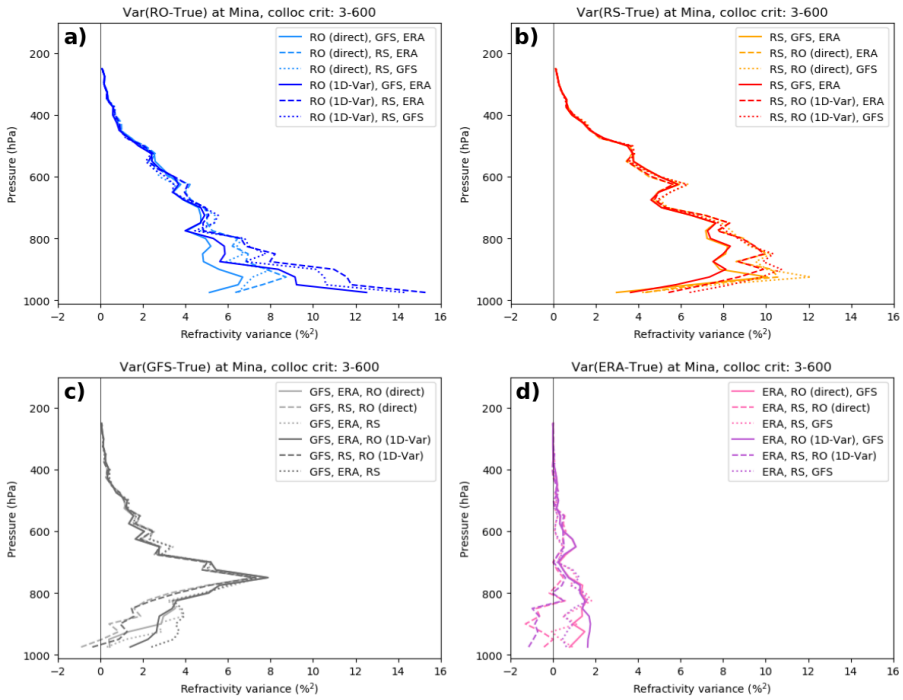

Figure 9. Estimated error variances (% squared) of refractivity at Mina: a) RO, b) RS, c) GFS, and d) ERA.

Figure 9 shows the estimates of the normalized refractivity errors for the five data sets. There is more spread in the refractivity estimates compared to those of the other variables, especially in the lower troposphere where the estimates vary between about 4 and 9 $\%^2$ (STD 2 to 3 %) for the two RO variances. Recall that the RO-Direct ~~N~~ N are the observed RO ~~N~~ N as provided by CDAAC while the RO 1D-VAR ~~N~~ N are modified based on the background (ERA) ~~N~~N. The average of the ~~N~~ N error variances for the radiosondes (~~Figure~~Fig. 9b) shows a maximum of $\sim$10 $\%^2$ (STD $\sim$3.2 %) around 900 hPa. The GFS error variance profiles show a maximum around 750 hPa of $\sim$8 $\%^2$ (STD $\sim$2.8 %). The ERA profiles show the smallest errors, with a maximum in the lower troposphere of an average of $\sim$2 $\%^2$ (STD $\sim$1.4 %). All data sets show a decrease of error variance to less than 0.5 $\%^2$ (STD$<$0.7 %) at 400 hPa. The reason for the large scatter in estimates of ~~N~~ N below about 800 hPa may be related to errors in ~~N~~ N caused by super-refraction in the lower troposphere, which occurs often in the tropics and subtropics. Super-refraction causes a negative ~~N~~ N bias, which may lead to larger error covariances in this layer. The smaller number of RO samples below 800 hPa (~~Figure~~Fig. 1) may also be a factor.

Figure 10 shows the mean of the ~~six~~ three estimates of the error variances of the five data sets for ~~q~~q, RH, ~~T and N~~ T, and N at Mina. The standard deviation[1] about these means is shown by the shaded areas. These figures show clearly the significant differences among the error variance estimates of the five data sets. In ~~Figure~~Fig. 10a, the error variance for specific humidity is greatest for the radiosonde (red and orange profiles) and least for the ERA profiles. As discussed earlier, the mean of the RO

[1] $\sigma = \left(\frac{1}{n-1}\sum_{n=1}^{n=3}(x_n - \bar{x})^2\right)^{1/2}$ where x_n denote the three error variance estimates

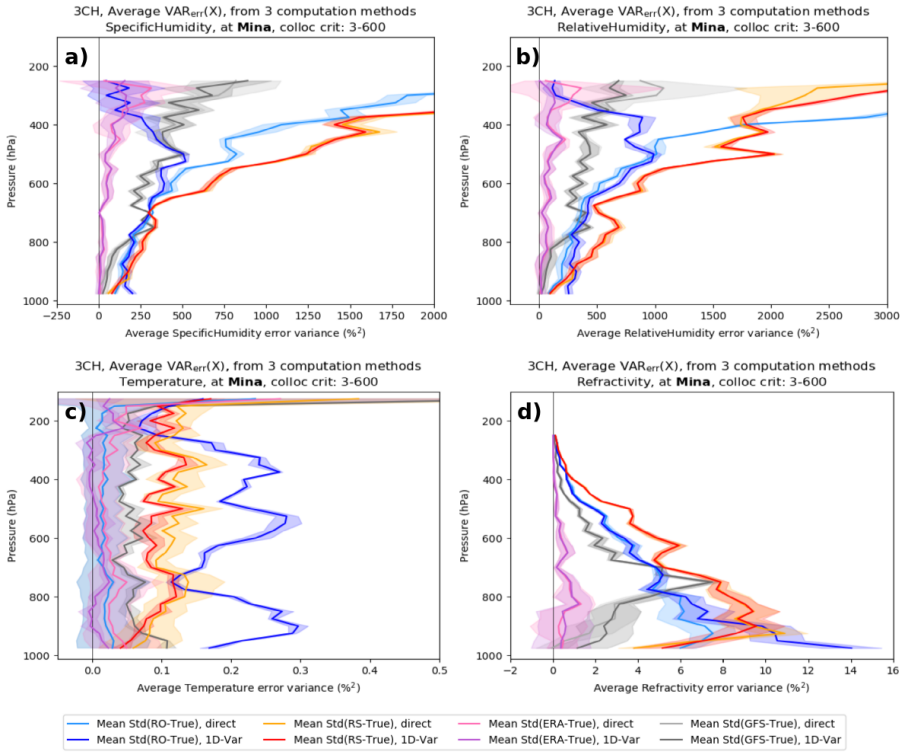

Figure 10. Mean of the ~~six~~ three estimates of error variance plots for ~~q~~*q*, RH, ~~T~~ *T* and ~~N~~ *N* using RO-Direct and RO 1D-VAR for each data set at Mina. The standard deviation about the mean is indicated by shaded areas. (a) specific humidity; (b) relative humidity; (c) temperature; and (d) refractivity. RO (blue), radiosonde (red), GFS (gray), and ERA (purple).

1D-VAR retrieval reaches a maximum at about 550 hPa and then decreases back toward zero as it becomes constrained by the background profile at high levels. Figures 10b–d show the mean profiles of error variance for relative humidity, temperature and refractivity. The relative humidity profiles are similar to the specific humidity profiles. The ERA errors are the smallest, followed by GFS, the RO and finally the radiosondes. The temperature error variance profiles show that the ERA errors are very close to zero throughout the entire troposphere. The GFS ~~and RO profiles are fairly constant with height at values of about~~ profiles (gray) and the RS profiles (red and orange) show relatively constant values with height of approximately 0.05 $\%^2$ and 0.1$\%^2$ respectively (~~0.22or~~ corresponding to temperature errors of 0.7 K ~~at 300K and 0.32or~~ and 0.9 K at 300 K~~respectively. The RS~~ , respectively). The RO shows an oscillating error variance profile ranging between 0.1 and 0.3 $\%^2$ (~~0.3and 0.5or~~ 0.9 K and ~~1.5~~1.6 K at 300 K~~respectively~~). Finally, the refractivity profiles show the greatest variability, but the mean profiles are still quite distinct. ERA again shows the lowest errors, followed by GFS, RO, and RS.

It is difficult to find previous results for RS temperature and specific humidity error variances. However, previous studies comparing RO with RS and models indicate that our estimates are reasonable and consistent with these studies. ~~Ho et al. (2017a)~~ Ho et al. (2 STD between RO and RS pairs for many RS types of about 1.5 K in the ~~layer~~ 200–20 hPa layer, where RO temperatures are most

Table 1. Normalized differences of zonal mean RO and ERA specific humidity in the tropics for cloudy conditions (computed from data in Vergados et al., 2014)

Pressure (hPa)	VAR ($\%^2$)	STD (%)
925	320	17.8
850	460	21.4
700	1260	35.5
500	2760	53.5
400	4220	65.0
300	5625	75.0

accurate (Table 2 in ~~Ho et al. (2017a)~~ Ho et al. (2017b)). This value corresponds to the apparent error between RS and RO, which is larger than the RS error. The estimated RS temperature error variances from ~~200 to 100~~200–100 hPa in ~~Figure~~Fig. 10c is about 0.15 $\%^2$, which corresponds to a STD of 0.39 % or 0.9 K for a mean temperature of 230 K. Ladstädter et al. (2015) compared high-quality ~~GRUAN~~ The GCOS (Global Climate Observing System) Reference Upper-Air Network (GRUAN) RS to RO globally and for a tropical station (Nauru) and subtropical station (Tateno, Japan) from 2002 to 2013. They found temperature STD of about 0.5 K for Nauru and 0.5 ~~-0.8~~to 0.8 K at Tateno averaged over the layer 800 ~~hPa~~ to 300 hPa. For specific humidity, they found STD between RO and RS of about 10 % increasing to about 40 % in the upper troposphere. In our calculations for Guam and the three subtropical Japanese stations our estimates for STD of q are similar (~~Figure~~Fig. 10a; and Appendix B ~~Figure~~Fig. B1), ranging from about 10 % at 900 hPa to 45 % at 300 hPa.

Ho et al. (2010) compared COSMIC RO observations to ECMWF analyses and several types of radiosondes for the period August–November, 2006. They found mean specific humidity STD of RO–ECMWF of $\sim$0.5 g kg^{-1} and RO–RS (Meisei) of $\sim$0.9 g kg^{-1}. From their plots of the vertical profiles of the STD, these numbers are typical for the layer 800–500 hPa, which, given the normalization values from the four RS stations in our study (~~Figure~~Fig. 3) of about 9 g kg^{-1} at 800 hPa and 2 g kg^{-1} at 500 hPa, correspond to ~~STD /VAR~~STD (VAR) values of $\sim$6 % ~~/~~(36 $\%^2$) at 800 hPa and 25 % ~~/~~(625 $\%^2$) at 500 hPa for ECMWF and $\sim$10 % ~~/~~(100 $\%^2$) at 800 hPa and 45 % ~~/~~(2025 $\%^2$) at 500 hPa for Meisei RS. These values are similar to the estimates of the ~~GFS analysis and~~ RS analysis for the Japanese stations shown in ~~Figure~~Fig. B1 of Appendix B.

Vergados et al. (2014) compared RO-derived observations of specific humidity with radiosondes and ERA-Interim under cloudy conditions in the tropics for August–October 2006. They used the direct method for computing specific humidity from the RO refractivity using the ERA-Interim temperatures. ~~They found the following~~ Their differences between zonal means of normalized RO and ERA-Interim observations of q are presented in Table 1 (we computed the normalized differences from their data in Table 3 for the tropics):

The VAR values in Table 1 correspond to apparent errors~~(RO compared to ERA~~ , where RO and ERA correspond to the observation and forecast variables respectively (Eq. (10)). As expected, they are larger than the estimated error variances for RO-Direct shown in ~~Figure~~Fig. 6a because ~~, as shown by Figure 5,~~ the apparent errors are always larger than the ~~estimated true~~

errors observation errors as shown in Eq. (4). This comparison indicates that the estimates of true errors in FigureFig. 6a are reasonable.

5.2 Summary of results at Naze, Ishi and Guam

The mean and STD error profiles for Naze, Ishi, and Guam corresponding to the above results for Mina are presented in Appendix B. Here we summarize the main similarities and differences between the error variance estimates for these stations compared to those for Mina. In general, we find similar magnitudes and shapes of the profiles of the estimated error variances of the five data sets for all four variables (qq, RH, T and NT, and N).

The estimated error profiles are especially similar for the three Japanese stations. This close similarity may be due primarily to the fact that the three locations are relatively close together and two of the three use the same type of radiosonde (Meisei).

The results from Guam are also similar in general magnitudes and shapes of the profiles to those from the three Japanese stations, but there are somewhat greater differences in some of the profiles (e.g. GFS q, RHand N and RS Nq, RH, and N; and RS N). These differences are likely due to the different location and the use of a different radiosonde type at Guam (VIZ/Sippican B2). The neglected error covariance terms are also likely different between the three Japanese stations, which are located in a data-rich region, and Guam, which is located in a data-sparse region. Thus the model errors are less likely to be highly correlated with a single observational system in the former than in the latter, where single observations may affect the models more significantly.

6 Summary and discussion

We used the "three cornered hat" (3CH) method to estimate vertical profiles of error variances of different observation and model data sets by computing the differences among the data sets at four fixed locations. We computed estimated error variances of four variables (specific humidity qq, relative humidity RH, temperature T and refractivity NT, and refractivity N) for five data sets (ERA, GFS, radiosondes (RS), RO-Direct and ROradio occultation (RO) 1D-VAR, and RO-Direct) at four different locations in the tropics and subtropics for the year 2007. The stations are Guam, Ishigakijima, Minamidaitojima, and Naze. The latter three stations are on Japanese islands and are located quite close together (a few hundred km kilometers apart). We computed vertical profiles of estimated error variances for normalized differences from the 2007 ERA mean values of qq, RH, T and N T and N at the four stations using three linearly independent equations (Eq. (7)–(9)) neglecting all error covariance terms(COV). Ideally, with a very large sample of data pairs and zero correlation of errors among the different data sets, all three equations would produce identical results. A However, a finite data set and , more importantly, non-zero error correlations among the data sets , lead to three different estimates, as shown by Rieckh and Anthes (2018) . The differences among the three estimates is a measure of these effects.

Although the neglect of the covariance terms affects the results to a noticeable degree in some of the estimated profiles, there is strong evidence that there is valid information in the estimated error profiles that rises above the noise caused by the neglect of the covariance terms and the limited data sample. This evidence is summarized as follows:

1. There is generally good agreement in the three estimated error profiles of the four variables for each of the five data sets at all four locations. It is unlikely that this agreement would occur by chance if the neglected error covariance terms were large enough to invalidate the results, because they would have to somehow combine or cancel in each of the three equations to give the observed similar results.

2. There are large differences in the overall structure (shape and magnitude) of the average vertical profiles of estimated error variances for the five data sets (~~Figure~~Fig. 10). These differences are significantly larger than the standard deviation ~~of the differences among the three methods of estimating~~ from the three independent equations used to compute the error variances.

3. The ~~vertical~~ variability, or ~~scatter~~ spread among the error estimates, is similar at most height levels for specific humidity, relative humidity, and temperature. If the error covariance terms were significant, they would almost certainly vary with height, giving different agreement in estimated error profiles with height. For example, we know that RO ~~is~~ temperature and refractivity are most accurate in the upper troposphere and least accurate in the lower troposphere ~~. Also,~~ and that the weight given to RO in the models' data assimilation varies significantly with height, being largest in the upper troposphere and smallest in the lower troposphere. Thus one would expect the RO-ERA and RO-GFS error covariance terms to vary significantly with height. Also, the RS errors as well as the ERA and GFS model errors vary with height. It is therefore unlikely that all of the neglected error covariance terms are the same at all heights.

4. The general structure and magnitudes of the estimated error variance profiles are similar at the four locations. However, there are some small differences among the profiles at the four locations. In general, the ~~vertical variability or scatter, which is~~ differences among the three estimates (indicated by the STD about the mean), which are a measure of the effect of the neglected covariance terms as well as limited sample size, ~~is~~ are smallest for Ishi, Naze, and Mina and largest for Guam. Since the three Japanese stations are close together, this suggests that there is a difference in the error variance of the RS observations at ~~these locations~~the Japanese RS observations compared to the Guam RS observations. There may also be small differences in the model errors over the Japanese stations, which are located in a data-rich area compared to Guam, which is located in a data-sparse region. The largest variability and largest error estimates occur at Guam, which uses a ~~different radiosonde , which~~ radiosonde that is thought to have large ~~dry and wet~~ water vapor biases due to sensor malfunctions (H. Vömel, personal communication, 2017).

5. The magnitudes of the estimated RO refractivity error variances are supported by previous published studies, including Kuo et al. (2004) and Chen et al. (2011).

6. The estimated ~~error profiles~~ errors are smallest for the ERA-Interim model data set, which is a reasonable result since ERA uses an excellent model and data assimilation system ~~, using~~ that assimilates many independent, quality checked observations. In fact, Vergados et al. (2015) state "ERA-Interim is one of the most advanced global atmospheric models simulating the state of the atmosphere with accuracy similar to what is theoretically possible (Simmons and Hollingsworth, 2002) using a 4D-Var method (Simmons et al., 2005)."

7. Our results show, in general, that the RO observations have smaller errors than the radiosonde errors, in agreement with previous studies.

Code availability. Code will be made available by the author upon request.

Data availability. Data can be made available from authors upon request.

Appendix A: Derivation of estimates of error variances using four data sets and the ~~*N*-cornered~~ N-cornered hat method

A1 Description of three-cornered hat method

In this appendix we summarize the "~~*N*-cornered hat"~~ three-cornered hat" (3CH) method (Gray and Allan, 1974) for estimating error variances from ~~*N* data sets. Variations and enhancements of the method have been applied to many diverse geophysical data sets, and for~~ three data sets~~it is called the "three-cornered hat" method (Wriley, 2003) or "triple collocation method" (Stoffelen, 1998)~~. Gray and Allan (1974) developed the method to estimate the absolute frequency stability of an ensemble of ~~*N*~~ N clocks by forming all (~~*N*-1)(*N*-2~~N-1)(N-2)/2 triads under the assumption that the clock errors are uncorrelated. Each of the triads are ~~the~~ three-cornered hat (~~THC~~3CH) estimates. ~~W.J. Wriley (2003)~~ Riley (2003) provides a summary of the ~~TCH methodand its history:~~ 3CH method.

~~The THC~~ Variations and enhancements of the 3CH method have been applied to many diverse geophysical data sets. The 3CH method has been used to estimate the stability of GNSS clocks using the measured frequencies from multiple clocks (Ekstrom and Koppang, 2006; Griggs et al., 2014, 2015; Luna et al., 2017). Valty et al. (2013) used the ~~TCH~~ 3CH method to estimate the geophysical load deformation computed from GRACE satellites, GPS vertical displacement measurements, and global general circulation (GCM) models. O'Carroll et al. (2008) compared three types of systems to measure sea-surface temperatures: two different radiometers and in situ observations from buoys. They discuss the assumption of the neglect of error correlations among the three data sets, the effect of representativeness errors, and the interpretation of "Truth", the true value of the variable being measured.

~~The "triple collocation method"~~ Closely related to the 3CH method is the triple co-location (TC) method, which was introduced by Stoffelen (1998), and has been widely used since in oceanography and hydrometeorology (e.g. Su et al., 2014; Gruber et al., 2016). It has been used to estimate the error variances of triplets of observation types to measure a diverse set of geophysical properties, including wave heights, sea surface temperatures, precipitation, surface winds over the ocean, leaf-area index products, and soil moisture. ~~For example,~~ Stoffelen (1998) estimated the error variances of in-situ measurements, ERS scatterometer winds, and NCEP (National Centers for Environmental Prediction) forecast model wind speeds. Later, ~~Vogelzang et al. (2011b)~~ Vogelzang et al. (2011a) compared four sets of scatterometer winds from ASCAT and Sea-

Winds with buoy measurements and ECMWF model forecasts of surface winds over the oceans to estimate the error variances and standard deviations of the different data sets and their combinations. Fang et al. (2012) ~~used Stoffelen's (1998) method to estimate~~ estimated the uncertainties in three different estimates of Leaf Area Index (LAI) products. ~~McColl et al. (2014)~~ McColl et al. (2014) extended the method by deriving a performance metric of the measurement system to the unknown truth, and applied the extended method to wind estimates from NWP, scatterometer and buoy wind estimates.

~~O'Carroll et al. (2008) compared three types of systems to measure sea-surface temperatures: two different radiometers and in situ observations from buoys. They discuss the assumption of the neglect of error correlations among the three data sets and the effect of representativeness errors.~~ Roebeling et al. (2012) used the triple ~~collocation~~ co-location method to estimate the errors associated with three ways of estimating precipitation: the Spinning Enhanced Visible and Infrared Imager (SEVERI), weather radars, and ground-based precipitation rain gauges. They concluded that the method provides useful error estimates of these systems.

The major assumption in the ~~above~~ 3CH and TC methods is that the errors of the three systems are uncorrelated. Correlations between any or all of the three measurement systems will reduce the accuracy of the error estimates. Other factors that can reduce the accuracy of the error estimates include widely different errors associated with the three systems or a small sample size. These factors can lead to negative estimates of error variances, especially when the estimates are close to zero (Gray and Allan, 1974; Riley, 2003) . All three of these factors potentially affect ~~our estimations here~~the 3CH estimates in this paper, but the general agreement of the three ~~methods~~ linearly independent equations for estimating the error variances of each variable suggests that the ~~estimations~~ estimates are still reasonably valid and contain useful information.

~~In this appendix~~

A2 Derivation of 3CH equations

n this section we summarize the derivation of the ~~N-Cornered hat~~ 3CH method as applied to four meteorological data sets, RO, RS, GFS and ERA. The error variance of a variable X (e.g. temperature, specific humidity, relative humidity, refractivity) is defined as ~~VAR(X)=$\sum$(X-X_t)^2/N where X_t~~

$$\mathrm{VAR_{err}(X)} = \frac{1}{n}\sum(\mathrm{X} - \mathrm{True})^2 = \frac{1}{n}\sum \mathrm{X_{err}^2} \tag{A1}$$

where True is the true ~~(but unknown)~~ but unknown value of X and the summation is over ~~N~~ n samples. Let RO correspond to ~~X_{RO} or~~ the value of X as estimated by RO, ERA correspond to ~~X_{ERA} or~~ the value of X as estimated by ERA, ~~True correspond to the true (but unknown) value of X,~~ and similarly for GFS and RS~~(radiosondes)~~. We then have

$$\mathrm{MS(RO-ERA) = VAR(RO) + VAR(ERA) - 2\,COV(RO, ERA)} \quad \mathrm{MS(RO-ERA) = VAR_{err}(RO) + VAR_{err}(ERA) - COV_{err}(RO, ERA)} \tag{A2}$$

where MS(RO – ERA) is the mean square difference between RO and ERA and the last term is the error covariance between RO and ERA.

In ~~our~~ the estimation of the ~~various error variances~~ error variances for the four data sets, we assume that the RO errors and ERA errors are uncorrelated, so the error covariance term in Eq. (A2) is zero, or in practice, negligibly small compared to the other terms). However, to show the complete (and exact) equations, we retain them here ~~for the six error variance~~ in the six equations involving the different ~~pair~~ pairs of data sets.

~~MS(RO – GFS) = VAR(RO) + VAR(GFS) – 2 COV(RO, GFS)~~MS(RO – GFS) = $\mathrm{VAR_{err}}$(RO) + $\mathrm{VAR_{err}}$(GFS) – 2 $\mathrm{COV_{err}}$(RO, GFS)

(A3)

~~MS(GFS – ERA) = VAR(GFS) + VAR(ERA) – 2 COV(GFS, ERA)~~MS(GFS – ERA) = $\mathrm{VAR_{err}}$(GFS) + $\mathrm{VAR_{err}}$(ERA) – 2 $\mathrm{COV_{err}}$(GFS, E

(A4)

~~MS(RO – RS) = VAR(RO) + VAR(RS) – 2 COV(RO, RS)~~MS(RO – RS) = $\mathrm{VAR_{err}}$(RO) + $\mathrm{VAR_{err}}$(RS) – 2 $\mathrm{COV_{err}}$(RO, RS)

(A5)

~~MS(RS – ERA) = VAR(RS) + VAR(ERA) – 2 COV(RS, ERA)~~MS(RS – ERA) = $\mathrm{VAR_{err}}$(RS) + $\mathrm{VAR_{err}}$(ERA) – 2 $\mathrm{COV_{err}}$(RS, ERA)

(A6)

~~MS(RS – GFS) = VAR(RS) + VAR(GFS) – 2 COV(RS, GFS)~~MS(RS – GFS) = $\mathrm{VAR_{err}}$(RS) + $\mathrm{VAR_{err}}$(GFS) – 2 $\mathrm{COV_{err}}$(RS, GFS)

(A7)

It is possible to use these six equations to get ~~six different~~three different, linearly independent estimates of the four unknowns ~~VAR(RO), VAR(ERA), VAR(GFS) and VAR(RS). For example, the full six equations for computing VAR(RO) are:~~ error variances for RO, RS, ERA, and GFS. For RO, these three $\mathrm{VAR_{err}}$ equations are:

~~2VAR(RO) =~~ $\mathrm{VAR_{err}}$(RO) =MS(RO – ERA) + MS(RO – GFS) – MS(GFS – ERA)]

+ ~~2 COV(RO, ERA) + COV(RO, GFS) – COV(GFS, ERA)~~2 [$\mathrm{COV_{err}}$(RO, ERA)+$\mathrm{COV_{err}}$(RO, GFS)–CO

(A8)

~~2VAR(RO) =~~ $\mathrm{VAR_{err}}$(RO) =MS(RO – ERA) + MS(RO – RS) – MS(RS – ERA)]

+ ~~2 COV(RO, ERA) + COV(RO, RS) – COV(RS, ERA)~~2 [$\mathrm{COV_{err}}$(RO, ERA)+$\mathrm{COV_{err}}$(RO, RS)–$\mathrm{COV_{err}}$

(A9)

~~2VAR(RO) =~~ $\mathrm{VAR_{err}}$(RO) =MS(RO – GFS) + MS(RO – RS) – MS(RS – GFS)]

+ ~~2 COV(RO, GFS) + COV(RO, RS) – COV(RS, GFS)~~2 [$\mathrm{COV_{err}}$(RO, GFS)+$\mathrm{COV_{err}}$(RO, RS)–$\mathrm{COV_{err}}$(

(A10)

O'Carroll et al. (2008) present these equations for a system of three observation types (their Eq. (1)). ~~Additionally, VAR(ERA), VAR(GFS), and VAR(RS) can be computed separately from different combinations of~~

As noted by an anonymous reviewer, it is possible to derive infinitely many linearly dependent equations by combining Eqs. (~~1~~A8)–(~~6), and these values can be substituted into Eqs. (1), (2), and (4) to compute the remaining three estimates of VAR(RO):~~

$$\mathrm{VAR(RO)} = \mathrm{MS(RO - ERA)} + 2\,\mathrm{COV(RO, ERA)} - \mathrm{VAR(ERA)} \tag{10}$$

$$\mathrm{VAR(RO)} = \mathrm{MS(RO - GFS)} + 2\,\mathrm{COV(RO, GFS)} - \mathrm{VAR(GFS)} \tag{11}$$

$$\mathrm{VAR(RO)} = \mathrm{MS(RO - RS)} + 2\,\mathrm{COV(RO, RS)} - \mathrm{VAR(RS)} \tag{12}$$

~~where VAR(ERA) is computed from Eqs~~A10) in different ways by forming combinations of the form $M_1 \times$Eq. (~~1), (2) and (3)~~

$$\begin{aligned}\mathrm{VAR(ERA)} = 1/2[&\mathrm{MS(GFS - ERA)} + \mathrm{MS(RS - GFS)} - \mathrm{MS(RS - GFS)}] \\ &+ \mathrm{COV(GFS, ERA)} + \mathrm{COV(RS, ERA)} - \mathrm{COV(RS, GFS)}\end{aligned} \tag{13}$$

~~VAR(GFS)is computed from Eqs~~A8)+$M_2 \times$Eq. (~~1), (4) and (5)~~

$$\begin{aligned}\mathrm{VAR(GFS)} = 1/2[&\mathrm{MS(GFS - ERA)} + \mathrm{MS(RS - GFS)} - \mathrm{MS(RS - ERA)}] \\ &+ \mathrm{COV(GFS, ERA)} + \mathrm{COV(RS, GFS)} - \mathrm{COV(RS, ERA)}\end{aligned} \tag{14}$$

~~and VAR(RS)is computed for Eqs~~A9)+$M_3 \times$Eq. (~~2), (4)~~ A10) where M_1, M_2 and, ~~and (6)~~

$$\begin{aligned}\mathrm{VAR(RS)} = 1/2[&\mathrm{MS(RS - ERA)} + \mathrm{MS(RS - GFS)} - \mathrm{MS(GFS - ERA)}] \\ &+ \mathrm{COV(RS, ERA)} + \mathrm{COV(RS, GFS)} - \mathrm{COV(GFS, ERA)}\end{aligned} \tag{15}$$

~~The same procedure can be used to derive six equations for estimating the error variances for the other three data sets, RS, ERA, and GFS (not shown).~~

~~So for each of the data sets RO, RS, ERA, and GFS, there are six different ways to estimate the respective error variances. The first three of these equations are linearly independent ; these are the three triads in the Gray and Allan (1974) N-cornered hat method, and the other three are linearly dependent, but are different ways of combining the observed data sets to estimate the error covariances~~M_3 are any numbers except those for which $M_1+M_2+M_3=0$. We did not pursue this possibility in this paper, but instead used the three linearly independent equations only in our estimates of error variances.

If all the neglected ~~error COV~~$\mathrm{COV}_{\mathrm{err}}$ terms were in fact identically zero ~~, the set of observations in the pairs (RO,ERA), (RO,GFS), (GFS,ERA), (RO,RS), (RS,ERA) and (RS,GFS) are the same (they are in our case),~~ and the sample size was very large (much larger than our sample size), all ~~six~~ three estimates of the error variances would be the same. The fact that they

give different solutions is because the neglected COV~~terms in are in realty~~ $_{err}$ terms are in reality not zero, and hence their neglect affects the ~~six~~ three approximate equations in different ways to give ~~six~~ three different solutions. The relatively small sample size ~~N~~ n also contributes to the differences in the ~~six~~ three solutions, which are a measure of these effects.

~~Figures ??–?? compare the mean of the estimated standard deviations of the errors (computed from the square root of the estimated error variances) associated with RO 1D-VAR, RS, ERA and GFS for specific humidity, relative humidity, temperature and refractivity computed from the six equations (left) and three equations (right) for Guam and Mina~~We also note that the error estimates contain any representativeness errors caused by the different data sets representing different scales of atmospheric structure (O'Carroll et al., 2008) . Representativeness errors can occur because of different horizontal or vertical resolutions or footprints of the data sets.

~~Mean of six and three error estimates of normalized specific humidity at Guam (top) and Mina (bottom). The standard deviation about the means is shown by the shading.~~

~~Mean of six and three error estimates of normalized relative humidity at Guam (top) and Mina (bottom). The standard deviation about the means is shown by the shading.~~

~~Mean of six and three error estimates of normalized temperature at Guam (top) and Mina (bottom). The standard deviation about the means is shown by the shading.~~

~~Mean of six and three error estimates of normalized refractivity at Guam (top) and Mina (bottom). The standard deviation about the means is shown by the shading.~~

~~There is generally good agreement among all the estimated error profiles of the four variables for each of the data sets at all four locations. The mean profiles are quite similar, but the scatter (STD) is smaller for the three independent solutions. It is unlikely that this agreement would occur by chance if the neglected error covariance terms were large enough to invalidate the results, because they would have to somehow combine or cancel in each of the three equations to give the observed similar results~~

A3 Brief comparison of 3CH method and triple co-location method

While it is not the intent of this paper to do a thorough comparison of the 3CH and triple co-location (TC) methods, which are introduced above, in response to a reviewer's comment we compared the two methods on a subset of our data sets. ~~The error covariance terms include error correlations between RO and ERA~~ The main difference between the 3CH and TC method is that the TC method corrects for additive and multiplicative biases among the three data sets, as discussed by Stoffelen (1998) , Vogelzang et al. (2011a) , and others. The TC method calibrates two of the data sets against the third, eliminating biases among the three data sets. As shown below, the TC method gave results very similar to the 3CH method for our data sets.

In our application of the TC method we use the following combinations of data sets: (ERA, RO and RS), (ERA, RO and GFS~~GFS and ERA RO and~~), and (ERA, GFS, RS). For the RO we use two RO retrievals, the Direct and 1D-VAR (see Sect. 2.3). The RO, RS and GFS data sets are all calibrated using ERA as the calibration reference, using the following calibration factors. For example, the calibrated RO and RS (designated by RO_{cal} and RS~~RS and ERA RS and GFS~~

~~These combine in different ways in the six equations; the neglected terms in each equation are : COV(RO, ERA) + COV(RO, GFS) – COV(GFS, ERA) COV(RO, ERA) + COV(RO, RS) – COV(RS, ERA) COV(RO, GFS) + COV(RO, RS) – COV(RS, GFS) 2 COV(RO, ERA) – COV(GFS, ERA) + COV(RS, ERA) – COV(RS, GFS) 2 COV(RO, GFS) – COV(GFS, ERA) + COV(RS, GFS) – COV(RS, ERA) 2 COV(RO, RS) – COV(RS, ERA) + COV(RS, GFS) – COV(GFS, ERA)~~ $_{\mathrm{cal}}$ respectively) using ERA as the reference are, following Stoffelen (1998) and Vogelzang et al. (2011a) :

$$\mathrm{RO}_{\mathrm{cal}} = \frac{\mathrm{RO} - b_{\mathrm{RO}}}{a_{\mathrm{RO}}}$$

$$\mathrm{RS}_{\mathrm{cal}} = \frac{\mathrm{RS} - b_{\mathrm{RS}}}{a_{\mathrm{RS}}}$$

where the additive bias terms are

$$b_{\mathrm{RO}} = \mathrm{M(RO - ERA)}$$

$$b_{\mathrm{RS}} = \mathrm{M(RS - ERA)}$$

and the multiplicative bias terms are

$$a_{\mathrm{RO}} = \frac{\mathrm{M(RO \cdot RS)}}{\mathrm{M(RS \cdot ERA)}}$$

$$a_{\mathrm{RS}} = \frac{\mathrm{M(RO \cdot RS)}}{\mathrm{M(RO \cdot ERA)}}$$

and M denotes the mean value over the data sets.

~~Since in all cases the results are similar, these six combinations must be approximately the same and most likely smaller that the terms involving the mean squares of the difference between the various data sets used to compute the estimates~~The results of the specific humidity error variance estimates for $\mathrm{RO}_{\mathrm{cal}}$ and $\mathrm{RS}_{\mathrm{cal}}$ compared to RO and RS are shown in Fig. A1 and the estimates for $\mathrm{ERA}_{\mathrm{cal}}$ and $\mathrm{GFS}_{\mathrm{cal}}$ compared to ERA and GFS are shown in Fig. A2. The left panels show the results from the TC method (calibrated data) and the right panels show the results using the 3CH method (uncalibrated data). The close similarity of the results indicates that the biases do not significantly affect the 3CH estimates, in agreement with the results from the error model study in Rieckh and Anthes (2018) .

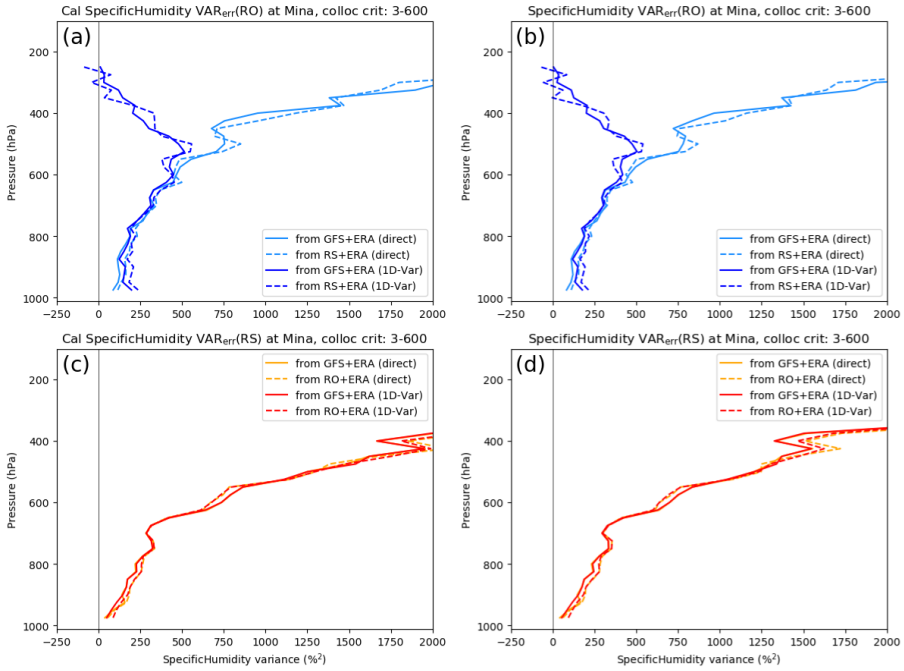

Figure A1. Estimated RO and RS error variances for specific humidity at Minamidaitojima (Japan) using calibrated data as in the TC method (left) and the uncalibrated data as in the 3CH method (right). For the TC method, the RO, RS, and GFS data sets are calibrated with respect to ERA as the reference data set. The following combinations of the 4 data sets are used: (ERA, RO, RS), (ERA, RO, GFS), and (ERA, GFS, RS).

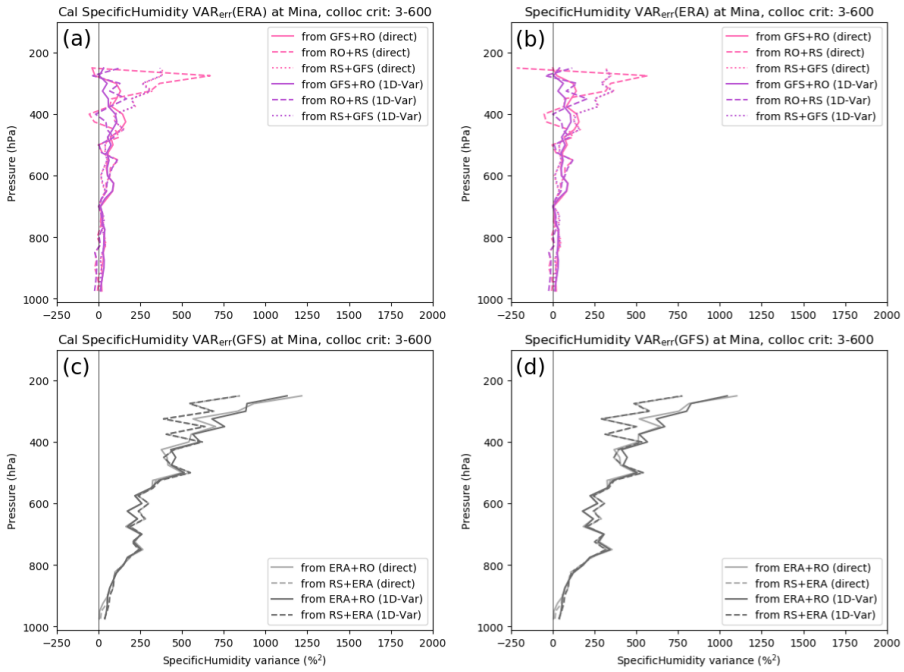

Figure A2. Estimated ERA and GFS specific error variances for ERA at Minamidaitojima (Japan) using the triple co-location (TC) method (left) and the three-cornered hat (THC) method (right). For the TC method, the RO, RS and GFS data sets are calibrated with respect to ERA. The following combinations of the 4 data sets are used: (ERA, RO, RS), (ERA, RO, GFS), and (ERA, GFS, RS).

Appendix B: Mean and standard deviations of three independent error estimates of ~~q~~*q*, RH, ~~T~~*T*, and ~~N~~ *N* using RO Direct and RO 1D-VAR at Guam, Ishi, Mina and Naze

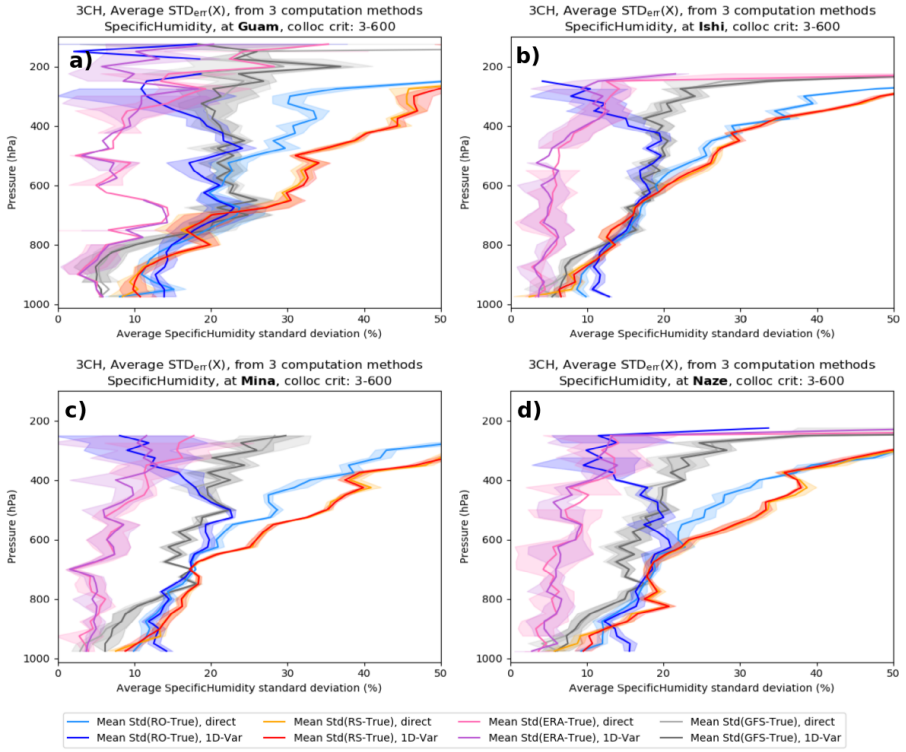

Figure B1. Mean and standard deviations (shading) of the three estimates of normalized specific humidity using RO Direct and RO 1D-VAR at (a) Guam, (b) Ishi, (c) Mina and (d) Naze.

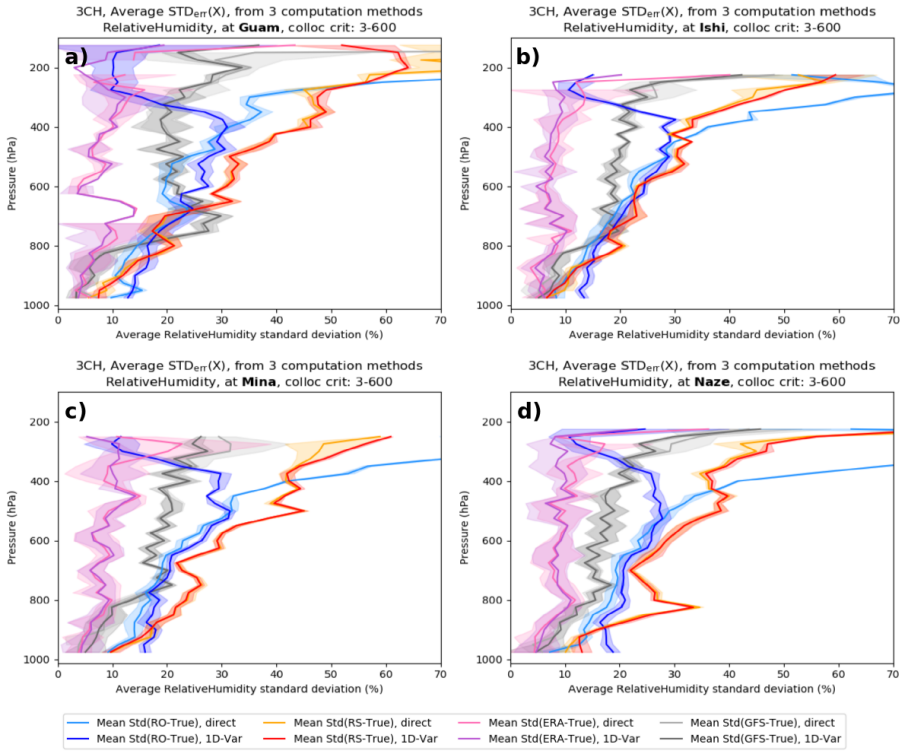

Figure B2. Mean and standard deviations (shading) of the three estimates of normalized ~~specific~~ relative humidity using RO Direct and RO 1D-VAR at (a) Guam, (b) Ishi, (c) Mina and (d) Naze.

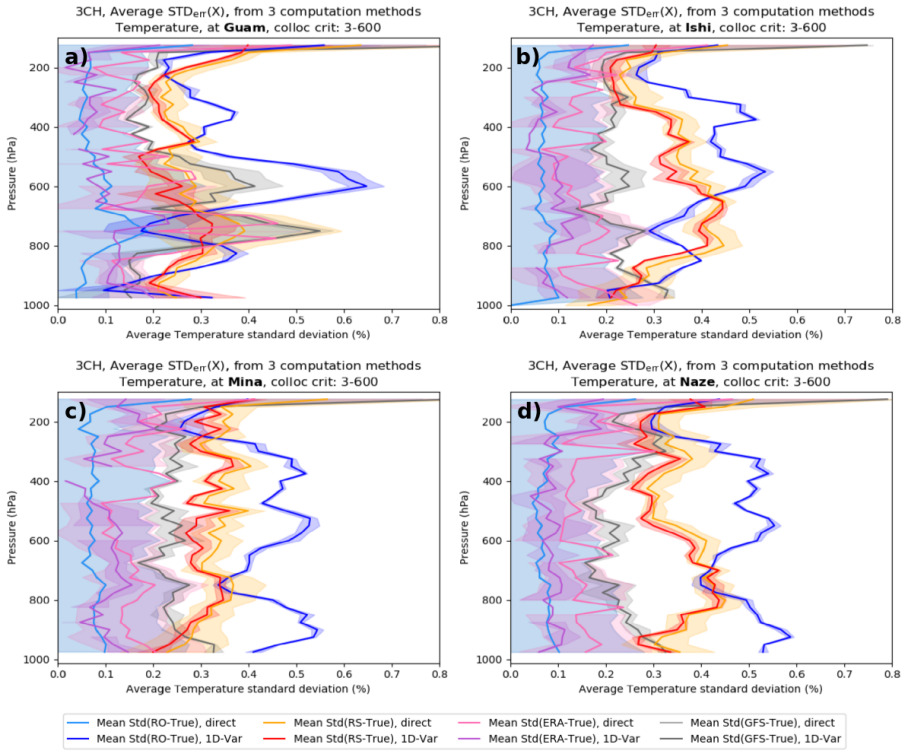

Figure B3. Mean and standard deviations (shading) of the three estimates of normalized ~~specific humidity~~ temperature using RO Direct and RO 1D-VAR at (a) Guam, (b) Ishi, (c) Mina and (d) Naze.

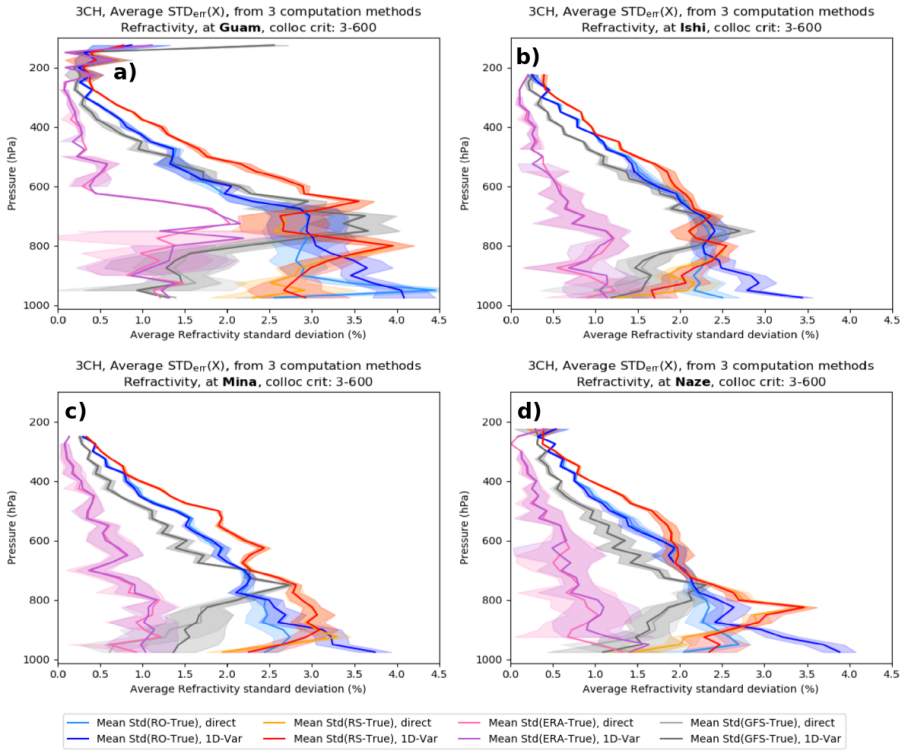

Figure B4. Mean and standard deviations (shading) of the three estimates of normalized ~~specific humidity~~ refractivity using RO Direct and RO 1D-VAR at (a) Guam, (b) Ishi, (c) Mina and (d) Naze.

Author contributions. R. Anthes formulated the overall idea of this work and T. Rieckh performed all the calculations and contributed significantly to the discussion of the results.

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

Acknowledgements. We acknowledge with thanks the insightful comments and advice on this study from Ian Culverwell (Met Office), Shay Gilpin (UCAR COSMIC), Sean Healy (ECMWF), Adrian Simmons (ECMWF) and Sergey Sokolovskiy (UCAR COSMIC). We thank the three anonymous reviewers for their constructive comments. Anthes and Rieckh were supported by NSF-NASA grant AGS-1522830. We thank Eric DeWeaver (NSF) and Jack Kaye (NASA) for their long-term support of COSMIC.

References

Chen, S.-Y., Huang, C.-Y., Kuo, Y.-H., and Sokolovskiy, S.: Observational Error Estimation of FORMOSAT-3/COSMIC GPS Radio Occultation Data, Mon. Wea. Rev., 139, 853–865, https://doi.org/10.1175/2010MWR3260.1, 2011.

Cucurull, L. and Derber, J. C.: Operational Implementation of COSMIC Observations into NCEP's Global Data Assimilation System, Wea. Forecasting, 23, 702–711, https://doi.org/10.1175/2008WAF2007070.1, 2008.

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., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, 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.

Desroziers, G. and Ivanov, S.: Diagnosing and adaptive tuning of observation-error parameters in a variational assimilation, Quart. J. Roy. Meteor. Soc., 127, 1433–1452, https://doi.org/10.1002/qj.49712757417, 2001.

Ekstrom, C. R. and Koppang, P. A.: Error Bars for Three-Cornered Hats, IEEE Trans. Ultrason. Ferroelect. Freq. Contr., 53, 876–879, https://doi.org/10.1109/TUFFC.2006.1632679, 2006.

Fang, H., Wei, S., Jiang, C., and Scipal, K.: Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method, Remote Sensing of Environment, 124, 610–621, https://doi.org/10.1016/j.rse.2012.06.013, 2012.

Gilpin, S., Rieckh, T., and Anthes, R.: Reducing representativeness and sampling errors in radio occultation–radiosonde comparisons, Atmos. Meas. Tech. Discuss., in preparation, 2017.

Gilpin, S., Rieckh, T., and Anthes, R.: Reducing representativeness and sampling errors in radio occultation–radiosonde comparisons, Atmos. Meas. Tech., 11, 2567–2582, https://doi.org/10.5194/amt-11-2567-2018, 2018.

Gray, J. E. and Allan, D. W.: A method for estimating the frequency stability of an individual oscillator, Atlantic City, New Jersey, May 29–31, 1974.

Griggs, E., Kursinski, E., and Akos, D.: An investigation of GNSS atomic clock behaviour at short time intervals, GPS Solut., 18, 443–452, https://doi.org/10.1007/s10291-013-0343-7, 2014.

Griggs, E., Kursinski, E., and Akos, D.: Short-term GNSS satellite clock stability, Radio Sci., 50, 813–826, https://doi.org/10.1002/2015RS005667, 2015.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, Int. J. Appl. Earth Obs. and Geoinf., 45, 200–211, https://doi.org/10.1016/j.jag.2015.09.002, 2016.

Ho, S.-P., Zhou, X., Kuo, Y.-H., Hunt, D., and Wang, J.-H.: 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., Mears, C., and Anthes, R. A.: Comparison of Global Observations and Trends of Total Precipitable Water Derived from Microwave Radiometers and COSMIC Radio Occultation from 2006 to 2013, ACPD, https://doi.org/10.5194/acp-2017-525, in review, 2017a.

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

Hollingsworth, A. and Lönnberg, P.: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, Tellus A, 38A, https://doi.org/10.1111/j.1600-0870.1986.tb00460.x, 1986.

Kleist, D. T., Parrish, D. F., Derber, J. C., Treadon, R., Wu, W.-S., and Lord, S.: Introduction of the GSI into the NCEP Global Data Assimilation System, Wea. Forecasting, 24, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1, 2009.

Kuo, Y.-H., Wee, T.-K., Sokolovskiy, S., Rocken, C., Schreiner, W., Hunt, D., and Anthes, R. A.: Inversion and error estimation of GPS radio occultation data, J. Meteor. Soc. Japan, 82, 507–531, 2004.

Ladstädter, F., Steiner, A. K., 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.

Luna, D., Pérez, D., Cifuentes, A., and Gómez, D.: Three-Cornered Hat Method via GPS Common-View Comparisons, IEEE Trans. Instrument. Measure., 66, 2143–2147, https://doi.org/10.1109/TIM.2017.2684918, 2017.

McColl, K., Vogelzang, J., Konings, A., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophys. Res. Lett., 41, 6229–6236, https://doi.org/10.1002/2014GL061322, 2014.

O'Carroll, A. G., Eyre, J. R., and Saunders, R. S.: Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations, J. Atmos. Oceanic Tech., 25, 1197–1207, https://doi.org/10.1175/2007JTECHO542.1, 2008.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's Spectral Statistical-Interpolation analysis system, Mon. Wea. Rev., 120, 1747–1763, https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2, 1992.

Rieckh, T. and Anthes, R.: Evaluating two methods of estimating error variance from multiple data sets using an error model, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-75, in review, 2018.

Rieckh, T., Anthes, R., Randel, W., Ho, S.-P., and Foelsche, U.: Tropospheric dry layers in the tropical western Pacific: comparisons of GPS radio occultation with multiple data sets, Atmos. Meas. Tech., 10, 1093–1110, https://doi.org/10.5194/amt-10-1093-2017, 2017.

Rieckh, T., Anthes, R., Randel, W., Ho, S.-P., and Foelsche, U.: Evaluating tropospheric humidity from GPS radio occultation, radiosonde, and AIRS from high-resolution time series, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2017-486, accepted, 2018.

Riley, W. J.: http://www.wriley.com/3-CornHat.htm, 2003.

Roebeling, R. A., Wolters, E. L. A., Meirink, J. F., and Leijnse, H.: Triple collocation of summer precipitation retrievals from SEVIRI over Europe with gridded rain gauge and weather radar data, Journal of Hydrometeorology, 13, 1552–1566, https://doi.org/10.1175/JHM-D-11-089.1, 2012.

Simmons, A. J. and Hollingsworth, A.: Some aspects of the improvement in skill of numerical prediction, Quart. J. Roy. Meteor. Soc., 128, 647–677, https://doi.org/10.1256/003590002321042135, 2002.

Simmons, A. J., Hortal, M., Kelly, G., McNally, A., Untch, A., and Uppala, S.: ECMWF Analyses and Forecasts of Stratospheric Winter Polar Vortex Breakup: September 2002 in the Southern Hemisphere and Related Events, J. Atmos. Sci., 62, 668–689, https://doi.org/10.1175/JAS-3322.1, 2005.

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

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, J. Geophys. Res., 103, 7755–7766, https://doi.org/10.1029/97JC03180, 1998.

Su, C.-H., Ryu, D., Crow, W. T., and Western, A. W.: Beyond triple collocation: Applications to soil moisture monitoring, J. Geophys. Res., 119, 6419–6439, https://doi.org/10.1002/2013JD021043, 2014.

Valty, P., de Viron, O., Panet, I., Camp, M. V., and Legrand, J.: Assessing the precision in loading estimates by geodetic techniques in Southern Europe, Geophys. J. Int., 194, 1441–1454, https://doi.org/https://doi.org/10.1093/gji/ggt173, 2013.

Vergados, P., Mannucci, A. J., and Ao, C. O.: 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., 119, 7718–7731, https://doi.org/10.1002/2013JD021398, 2014.

Vergados, P., Mannucci, A. J., Ao, C. O., Jiang, J., and Su, H.: On the comparisons of tropical relative humidity in the lower and middle troposphere among COSMIC radio occultations and MERRA and ECMWF data sets, Atmos. Meas. Tech., 8, 1789–1797, https://doi.org/10.5194/amt-8-1789-2015, 2015.

Vogelzang, J., Stoffelen, A., Verhoef, A., and Figa-Saldaña, J.: On the quality of high-resolution scatterometer winds, J. Geophys. Res., 116, https://doi.org/10.1029/2010JC006640, 2011a.

Vogelzang, J., Stoffelen, A., Verhoef, A., and Figa-Saldaña, J.: On the quality of high-resolution scatterometer winds, J. Geophys. Res., 116, https://doi.org/10.1029/2010JC006640, 2011b.

Wriley, W. J.: http://www.wriley.com/3-CornHat.htm, 2003.