Estimating Error Variances of a Microwave Sensor and Dropsondes aboard the Global Hawk in Hurricanes Using the Three-Cornered Hat Method

ANDREW C. KREN[a,b] AND RICHARD A. ANTHES[c]

[a] *Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida*
[b] *NOAA/Atlantic Oceanographic and Meteorological Laboratory/Hurricane Research Division, Miami, Florida*
[c] *COSMIC Program Office, University Corporation for Atmospheric Research, Boulder, Colorado*

(Manuscript received 20 March 2020, in final form 29 September 2020)

ABSTRACT: This study estimates the random error variances and standard deviations (STDs) for four datasets: Global Hawk (GH) dropsondes (DROP), the High-Altitude Monolithic Microwave Integrated Circuit Sounding Radiometer (HAMSR) aboard the GH, the fifth European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5), and the Hurricane Weather Research and Forecasting (HWRF) Model, using the three-cornered hat (3CH) method. These estimates are made during the 2016 Sensing Hazards with Operational Unmanned Technology (SHOUT) season in the environment of four tropical cyclones from August to October. For temperature and specific and relative humidity, the ERA5, HWRF, and DROP datasets all have similar magnitudes of errors, with ERA5 having the smallest. The error STDs of temperature and specific humidity are less than 0.8 K and 1.0 g kg^{-1} over most of the troposphere, while relative humidity error STDs increase from less than 5% near the surface to between 10% and 20% in the upper troposphere. The HAMSR bias-corrected data have larger errors, with estimated error STDs of temperature and specific humidity in the lower troposphere between 1.5 and 2.0 K and between 1.5 and 2.5 g kg^{-1}. HAMSR's relative humidity error STD increases from approximately 10% in the lower troposphere to 30% in the upper troposphere. The 3CH method error estimates are generally consistent with prior independent estimates of errors and uncertainties for the HAMSR and dropsonde datasets, although they are somewhat larger, likely due to the inclusion of representativeness errors (differences associated with different spatial and temporal scales represented by the data).

KEYWORDS: Tropical cyclones; Aircraft observations; Dropsondes; Remote sensing; Error analysis; Reanalysis data

1. Introduction

NOAA's Sensing Hazards with Operational Unmanned Technology (SHOUT; Dunion et al. 2018; Wick et al. 2018a, 2020) program during 2015–16 was designed partially to evaluate the effectiveness of the NASA unmanned aerial systems (UAS) Global Hawk (GH) on tropical cyclone (TC) prediction. Several sensitivity studies since this campaign have evaluated the impact of the high-altitude dropsonde observations, showing added value for both track and intensity metrics in global and regional modeling systems (Christophersen et al. 2017, 2018a,b; Kren et al. 2018). Additional studies are underway to obtain more robust statistics on the overall value of the GH dropsondes (Wick et al. 2020). These periods include years when SHOUT partnered with other field campaigns targeting TCs: flights in 2012–14 during the Hurricane and Severe Storm Sentinel (HS3; Braun et al. 2016) campaign and in 2017 with the Eastern Pacific Origins and Characteristics of Hurricanes (EPOCH; Emory et al. 2015) program.

In addition to dropsondes, the GH provides a suite of remote sensors to probe the hurricane environment and TC core. One of these sensors is the High-Altitude Monolithic Microwave Integrated Circuit Sounding Radiometer (HAMSR; Brown et al. 2011). HAMSR measures microwave radiances over 25 spectral channels at a high temporal (1.1 s) resolution and a ~2 km vertical resolution. These radiances may be used to retrieve temperature and water vapor profiles in the clear or cloudy TC environment. A few preliminary studies have examined the impact of using HAMSR data in numerical weather prediction (NWP) models (Wick et al. 2020). However, the impact has been mixed, and the assimilation of the HAMSR retrievals is likely far from optimal.

Knowing the error characteristics, including the uncertainty, of any observational or model dataset is of critical scientific importance. Estimating the uncertainty, or random errors, is also important for applications of any dataset, including risk assessment and decision making. Thus, users of observational or model datasets should be aware of the errors in the data, while developers of observational systems or models need to know the error characteristics of their data in order to validate and improve their products (reduce the errors). Correct error characterization is vital for proper data assimilation into NWP models (Desroziers and Ivanov 2001).

Current work is underway to assimilate HAMSR temperature and moisture retrievals into the Hurricane and Weather Research and Forecasting (HWRF) Model as part of a SHOUT follow-on study to make better use of the remote sensing data aboard the GH. To our knowledge, no prior work has been done to accurately characterize the error variance of the HAMSR retrievals. In addition, a thorough examination of the errors of dropsondes released from the GH has yet to be carried out (Wick et al. 2018b).

This paper estimates the errors of the GH dropsonde and HAMSR datasets, as well as two model datasets (HWRF and ERA5) during the SHOUT campaign using the

Corresponding author: Andrew Charles Kren, andrew.kren@noaa.gov

DOI: 10.1175/JTECH-D-20-0044.1

"three-cornered hat" (3CH) method (Gray and Allan 1974; Anthes and Rieckh 2018; Sjoberg et al. 2021). In the 3CH method, random error variances and standard deviations (STDs) of three or more independent collocated datasets are estimated simultaneously by forming differences between combinations of the three datasets, which can be models or observations. An important requirement for the accuracy of the 3CH method is that the random errors of the datasets be uncorrelated, or have small correlations (Rieckh and Anthes 2018; Sjoberg et al. 2021). Brief and preliminary results from this study were reported in Wick et al. (2020).

Section 2 summarizes the four datasets used in the error estimations. Section 3 summarizes the 3CH method and section 4 compares the 3CH error estimates with other error estimates. Section 5 presents the conclusions.

2. Datasets

We estimate the errors of the four datasets from August through October 2016 in the environment of TCs. The year 2016 is chosen due to the prevalence of GH dropsonde deployments and HAMSR retrievals during SHOUT. Four TCs were sampled during this period over the Atlantic Ocean: Gaston, Hermine, Karl, and Matthew. The HAMSR retrievals in 2016 were bias corrected by comparing to GH dropsondes [S. Brown, NASA's Jet Propulsion Laboratory (JPL), 2020, personal communication] and thus are of higher quality than in prior flight missions.

a. ERA5

ERA5 is the latest global reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF; Hersbach et al. 2018). It replaces ERA-Interim and contains improvements relative to past reanalyses. The current period of ERA5 runs from 1979 to the present and contains hourly estimates of atmospheric, land, and ocean variables at a 31-km horizontal resolution. In the vertical there are 137 hybrid sigma levels from the surface to 80 km (0.01 hPa), with atmospheric variables interpolated to 37 pressure levels. While the resolution of ERA5 is not sufficient to properly describe the inner region of TCs, it is used as an additional dataset to characterize HAMSR errors.

ERA5 assimilates many types of conventional and satellite observations, but it does not assimilate HAMSR observations, so any error correlations between ERA5 and HAMSR is likely small. However, ERA5 does assimilate GH dropsondes and thus there may be some correlation between ERA5 and dropsonde errors. However, there are many more observations going into the analysis than this one dataset alone, so that the correlations of ERA5 and dropsonde errors is assumed to be small.

b. HWRF Model

HWRF is a nonhydrostatic, coupled atmospheric–ocean model, employing the dynamical core of the Weather Research and Forecasting (WRF) mesoscale model (Gopalakrishnan et al. 2011; Tallapragada et al. 2014). It contains one parent domain with size of 77.2° × 77.2° centered on the TC initial position, and two smaller nested domains, with a horizontal resolution of the three grids at 13.5, 4.5, and 1.5 km. HWRF contains output on 46 pressure levels from 1000 to 2 hPa. This study uses HWRF analyses for the 2016 period that were generated from the 2017 version of HWRF, produced at the National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC). The HWRF analyses used in this study are based on the NCEP Global Forecast System analyses, interpolated to the HWRF outer grid (13.5 km resolution); data assimilation is also performed on the two nested domains to produce an updated analysis at each time step.

HWRF and ERA5 are two different models with varying resolution and physics. They both assimilate dropsonde observations from the GH, but HWRF does not assimilate all observations (in both number of observations and types) as in ERA5. Another difference between HWRF and ERA5 is that HWRF assimilates inner-core dropsondes and assimilates these at the correct location using dropsonde drift (Aberson et al. 2017). This dropsonde drift correction in HWRF was not operational for the 2016 cases discussed herein and thus HWRF only assimilated the dropsonde data at a fixed location. The correlation between the errors in these datasets is likely to be small. However, HWRF errors may be correlated with the dropsonde errors.

c. GH dropsondes

The dropsondes (DROP) released from the NASA GH measure pressure, temperature, and relative humidity at 2 Hz, as well as wind speed and direction at 4 Hz. The temporal resolution roughly corresponds to a vertical resolution of 6 m for pressure, temperature, and humidity and 3 m for wind speed and direction (Wick et al. 2018b). The GH dropsonde observations were retrieved from https://www.esrl.noaa.gov/psd/psd2/coastal/satres/data/static/shout/2016_HRR.html. The dropsonde observations used here are the full resolution data, as opposed to the Binary Universal Form for the Representation of Meteorological Data (BUFR) format, which includes simply mandatory and significant levels assimilated in global and regional models (Aberson et al. 2017). Prior to the start of the 2016 hurricane season, a systematic dry bias was found in the GH dropsonde moisture observations. This bias was corrected prior to the four hurricanes that were sampled in this study (G. Wick 2019, personal communication). Last, the dropsonde drift is not accounted for in the collocation procedure discussed in section 2e. The maximum drift is estimated to be less than 15 km (Chan et al. 2018) and thus the error introduced by its neglect in the collocation is expected to be negligible.

d. HAMSR retrievals

The HAMSR cross-track scanning atmospheric sounder designed at JPL contains 8 sounding channels near the 60-GHz oxygen line complex, 10 channels near the 118.75-GHz oxygen line, and 7 near the 183.31-GHz water vapor line (Brown et al. 2011). Retrieval products of temperature and absolute humidity profiles are obtained using a one-dimensional plane-parallel radiative-transfer model. The first guess, or background, is determined using radiosonde profiles that are near the GH flight paths, as well as climatology (Brown et al. 2007). The resultant

retrievals are generated at a horizontal resolution of ~2 km (from an altitude of 20 km at nadir) and output every 1.1 s. The cross-track swath spans 45° from nadir, a roughly 45 km swath width at 20 km altitude. The HAMSR data used the GH dropsonde data in a bias-correction process, using coincident dropsondes released from drone aircraft. The average difference between HAMSR and the dropsondes from all flights (not only the ones used in this study) was used to remove the bias in HAMSR retrievals [S. Brown, NASA's JPL, 2020, personal communication]. Since this bias correction procedure possibly introduced some correlation of errors between the dropsonde and HAMSR data, we also estimated the errors of the HAMSR retrievals without bias corrections (denoted HAMSR-NC). This uncorrected HAMSR dataset contained a few profiles which appeared to contain large errors (outliers), and so these were eliminated in our 3CH estimates. For these two HAMSR datasets, we included all scan angles. We also estimated the STD of errors of the uncorrected HAMSR data, but restricted the sample to those with scan angles less than 30° (denoted HAMSR-30), which are generally more accurate than those with larger scan angles (Brown et al. 2011). The variation in horizontal location of the HAMSR soundings with altitude for off-nadir scan angles is neglected in the collocation procedure. For a 30° scan angle, the maximum displacement from 16 km altitude is 8 km, and the error due to neglect of this displacement is expected to be negligible.

Neither ERA5 nor HWRF assimilates HAMSR operationally, making any correlation of errors between these datasets small. There may be some correlation of errors between the HAMSR bias-corrected retrievals and DROP. The HAMSR uncorrected retrievals were downloaded from https://www.esrl.noaa.gov/psd/psd2/coastal/satres/data/static/shout/2016_HRR.html, while the HAMSR bias-corrected data were received directly from Shannon Brown at JPL during the SHOUT campaign.

e. *Colocation of the datasets*

The 3CH method requires that all datasets be collocated to the same time and location; here we collocated the ERA5, HWRF, and HAMSR datasets to the dropsonde locations and times at release. Figure 1 shows the dropsonde release locations during the SHOUT 2016 season. There were a total of 634 dropsondes over the nine flight missions across the storms Gaston, Hermine, Karl, and Matthew. HAMSR retrievals are chosen that are within 60 min of the dropsonde release time. If vertical profiles are within this time threshold, we iterate through the subset (irrespective of scan angle) and select the profile that is closest in time and location to the dropsonde release location and time. Figure 2 shows that the majority of HAMSR profiles that satisfy these criteria are within 5 min and 5 km of the dropsonde time and position. The reason for varying distances of the HAMSR retrievals and the dropsonde release locations and times (Fig. 2) is because HAMSR is not continuously scanning at all times. In addition, the bias corrected retrievals likely filtered out outliers from the raw dataset, which led to fewer observations in close proximity to the dropsonde release times and locations. Of the 634 dropsondes

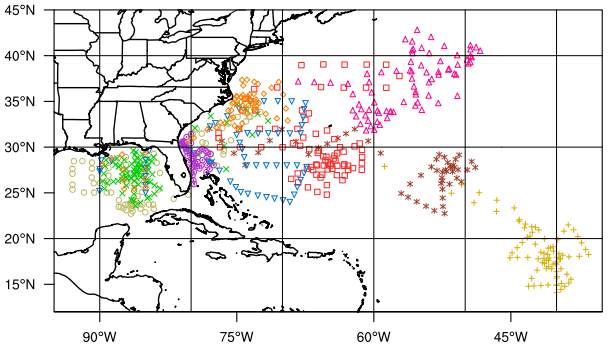

FIG. 1. Dropsonde release locations during the SHOUT 2016 season. The different colors (marker styles) denote different flight missions over the four storms sampled (nine total).

released in or near the four TCs (Fig. 1), we found 533 collocated datasets using the bias-corrected HAMSR data. When using the uncorrected HAMSR data at all scan angles, the elimination of 17 profiles with outliers reduced the collocated sample size to 516, while restricting the HAMSR profiles to those with scan angles less than 30° further reduced the sample size to 513.

To determine where the collocated data are positioned in relation to the TC centers, we interpolate the ERA5 sea level pressure (SLP) to the time of the dropsonde release for all collocations. The National Hurricane Center (NHC) best track dataset (HURDAT2) of each tropical cyclone is then used to match the best track date that is closest to the dropsonde release date. This date gives an initial estimate of the position of the TC center at each dropsonde release. Using the ERA5 interpolated SLP field, then, we find the minimum SLP within ±15° latitude and longitude of the best track estimated position to designate the storm center. The great circle distance between the storm center and the dropsonde location denotes the distance from the TC center. The reason we use a combination of ERA5 and HURDAT2 is because the dropsonde release times do not always coincide with the best track 6-houly analyses.

Figure 3a shows a histogram of the distance between the dropsonde locations and the TC centers. There is spread in the location of the dropsondes relative to the storm center, with several sampling the larger environment over the Atlantic Ocean. The vast majority of the dropsondes, however, are within 400 km of the TC center. Figure 3b shows an example of an ERA5 interpolated SLP field, valid at the dropsonde time of 0617 UTC 27 August 2016 during Gaston. The position of the dropsonde is denoted by the red circle in the northeast quadrant of Gaston.

Once the dropsondes and HAMSR retrievals are collocated, ERA5 temperature, specific and relative humidity are interpolated linearly in time and bilinearly in space to the dropsonde release times and locations. The dropsondes and ERA5 reanalysis are vertically interpolated to the 25 pressure levels of the HAMSR data: 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, and 1000 hPa. The HWRF analysis products, which

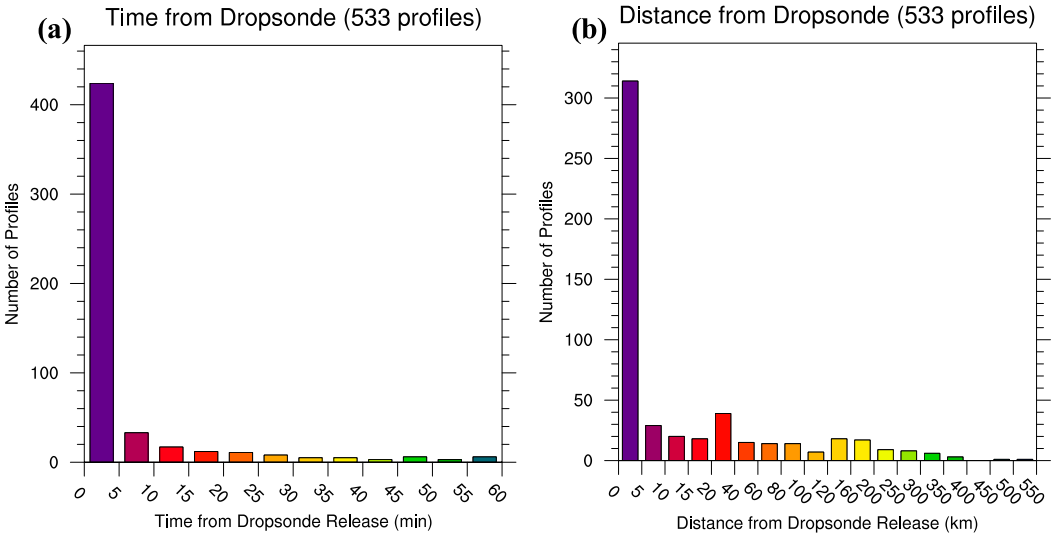

FIG. 2. Histogram of the (a) time (min) and (b) distance (km) separation of HAMSR retrieval profiles from the dropsonde release time and location.

are provided every 6 h, are also interpolated linearly in time and bilinearly in space to the dropsonde times and locations. The variables in HWRF exist at the pressure levels of the HAMSR data, so no vertical interpolation is needed. Since the dropsondes and HAMSR do not contain specific humidity directly, this variable is derived from the corresponding temperature, relative humidity, and pressure observations. Specific and relative humidities are provided in both the ERA5 and HWRF datasets.

To reduce sampling errors, a double-differencing spatial–temporal correction, as described in Gilpin et al. (2018), is applied to the HAMSR profiles. For example, the spatial–temporal corrected difference X^{SC} of a variable X between the HAMSR and DROP observations is computed as

$$X^{SC} = X_{HAMSR} - X_{DROP} - (X^{ERA5}_{HAMSR} - X^{ERA5}_{DROP}), \qquad (1)$$

where X_{HAMSR} is the HAMSR profile closest in space and time to DROP, X^{ERA5}_{HAMSR} is the ERA5 profile interpolated to the HAMSR time and location, and X^{ERA5}_{DROP} is the ERA5 profile interpolated in space and time to the dropsonde.

We compute the estimated errors of the 533 collocated datasets as well as the errors of the datasets normalized by the ERA5 mean profiles (averaged over the sample), similar to

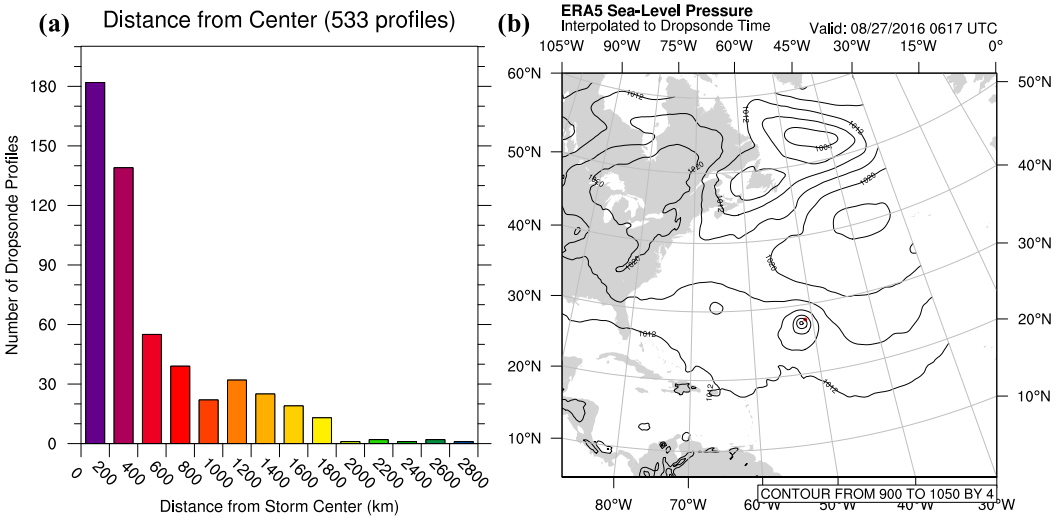

FIG. 3. (a) Histogram of the distance between the dropsonde release location and the tropical cyclone storm center using ERA5 sea level pressure (SLP) observations and NHC best track data. (b) Example ERA5 SLP interpolated to the dropsonde time for 0617 UTC 27 Aug 2016. The location of the dropsonde is shown in the red dot for the Hurricane Gaston case.

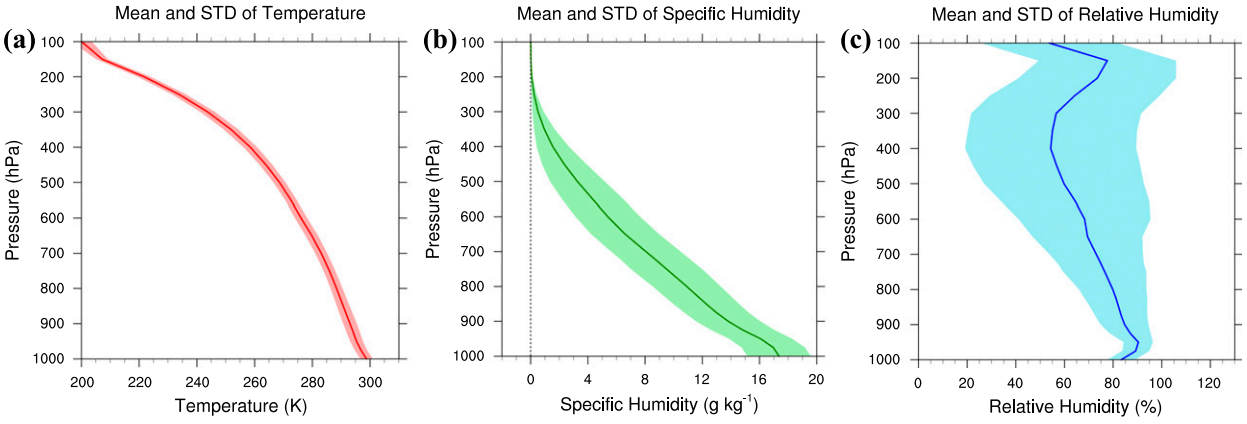

FIG. 4. ERA5 mean (solid line) and standard deviation (shading) of the 533 collocated profiles for (a) temperature (K), (b) specific humidity (g kg^{-1}), and (c) relative humidity (%). The mean profiles are used in the normalizations of the differences between datasets in the 3CH calculations.

Anthes and Rieckh (2018). These normalized error estimates are comparable to the observation minus background statistics commonly used by NWP centers. Figure 4 shows the mean ERA5 profiles for the three variables. The greatest variability in specific humidity is found below 300 hPa. For relative humidity, the largest spread exists in the upper-levels between 200 and 600 hPa. Temperature is much less variable. Figure 5 shows one example of temperature, specific and relative humidity for the four datasets corresponding to the dropsonde release at 0031 UTC 27 August 2016.

3. Estimation of error variances and standard deviations

Once we have the collocated differences between the datasets, the methodology to estimate the error variance of the four datasets closely follows that of Anthes and Rieckh (2018) and Sjoberg et al. (2021). The error variance of a dataset X is defined by

$$\mathrm{VAR}_{\mathrm{err}}(X) = \langle (X - X_{\mathrm{True}} - b_x)^2 \rangle, \tag{2}$$

where X_{True} is the true (but unknown) value of X, b_x is the bias of X with respect to X_{True}, and the brackets denote the sample mean. In the 3CH method, we use three linearly independent equations to estimate the error variance of each dataset by assuming that the error covariances among the datasets are negligible compared to the mean-square (MS) differences between the datasets. For example, the three independent estimates for the error variance for temperature, specific and relative humidity, excluding the neglected error covariance terms, for HAMSR are

$$\mathrm{VAR}_{\mathrm{err}}(\mathrm{HAMSR}) = \frac{1}{2}\mathrm{MS}(\mathrm{HAMSR} - \mathrm{ERA5}) + \frac{1}{2}\mathrm{MS}(\mathrm{HAMSR} - \mathrm{DROP}) - \frac{1}{2}\mathrm{MS}(\mathrm{ERA5} - \mathrm{DROP}) - \frac{1}{2}[b^2_{\mathrm{HAMSR,ERA5}} + b^2_{\mathrm{HAMSR,DROP}} - b^2_{\mathrm{ERA5,DROP}}], \tag{3}$$

$$\mathrm{VAR}_{\mathrm{err}}(\mathrm{HAMSR}) = \frac{1}{2}\mathrm{MS}(\mathrm{HAMSR} - \mathrm{ERA5}) + \frac{1}{2}\mathrm{MS}(\mathrm{HAMSR} - \mathrm{HWRF}) - \frac{1}{2}\mathrm{MS}(\mathrm{ERA5} - \mathrm{HWRF}) - \frac{1}{2}[b^2_{\mathrm{HAMSR,ERA5}} + b^2_{\mathrm{HAMSR,HWRF}} - b^2_{\mathrm{ERA5,HWRF}}], \tag{4}$$

$$\mathrm{VAR}_{\mathrm{err}}(\mathrm{HAMSR}) = \frac{1}{2}\mathrm{MS}(\mathrm{HAMSR} - \mathrm{DROP}) + \frac{1}{2}\mathrm{MS}(\mathrm{HAMSR} - \mathrm{HWRF}) - \frac{1}{2}\mathrm{MS}(\mathrm{DROP} - \mathrm{HWRF}) - \frac{1}{2}[b^2_{\mathrm{HAMSR,DROP}} + b^2_{\mathrm{HAMSR,HWRF}} - b^2_{\mathrm{DROP,HWRF}}], \tag{5}$$

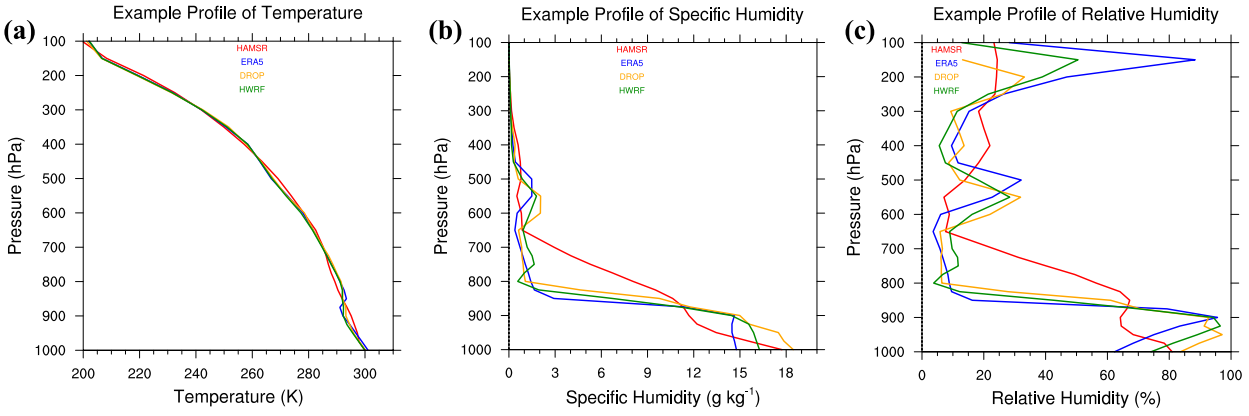

FIG. 5. Sample profiles of (a) temperature (K), (b) specific humidity (g kg^{-1}), and (c) relative humidity (%) for the four datasets: ERA5 (blue), HWRF (green), HAMSR (red), and DROP (orange). All profiles are collocated to the dropsonde location release at 0031 UTC 27 Aug 2016.

where $\mathrm{VAR}_{\mathrm{err}}(\mathrm{HAMSR})$ denotes the error variance of HAMSR, MS denotes the mean square difference of the two datasets, and the b terms denote the mean biases between any two datasets X and Y,

$$b_{x,y} = \langle X - Y \rangle, \tag{6}$$

where the brackets denote the sample mean. Similar equations are derived for the estimated error variances of ERA5, DROP, and HWRF.

In addition to calculating the error variance, we also calculated the root-mean-square (RMS) differences,

$$\sqrt{\mathrm{MS}(X - Y)}, \tag{7}$$

and the variance and STD of the differences between the datasets,

$$\mathrm{VAR}(X - Y) = \mathrm{MS}(X - Y)^2 - b_{x,y}^2, \tag{8}$$

$$\mathrm{STD}(X - Y) = \sqrt{\mathrm{VAR}(X - Y)}. \tag{9}$$

The STD of the estimated errors is the square root of the estimated error variance of a variable X. Last, the means and STDs of the three error variance estimates for each dataset are calculated from

$$\sigma = \left[\frac{1}{2}\sum_{n=1}^{n=3}(x_n - \bar{x})^2\right]^{1/2}, \tag{10}$$

where x_n denotes the nth error variance estimates and $\bar{x}$ is the mean of the three estimates.

4. Results

a. *Mean and STD of differences from ERA5*

Before showing the estimated 3CH errors, Fig. 6 shows the mean and STD of the differences of HAMSR, HWRF, and DROP from ERA5 for temperature and specific and relative

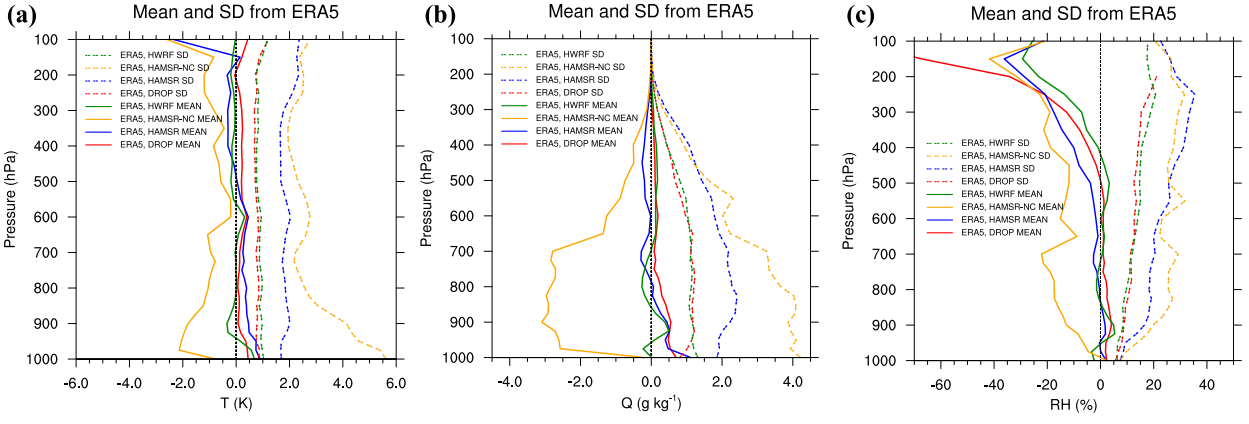

FIG. 6. Mean (solid lines) and standard deviation (STD; dashed lines) of differences between HWRF (green), DROP (red), HAMSR (blue), and HAMSR-NC (all scan angles, not bias corrected; orange) and ERA5 for (a) temperature (K), (b) specific humidity (g kg^{-1}), and (c) relative humidity (%).

humidity. Results indicate the bias and spread from the ERA5 profiles.

For temperature (Fig. 6a) over most of the troposphere, the mean differences from ERA5 are within ±1 K for the HWRF, DROP, and bias-corrected HAMSR data. The mean differences of DROP and ERA5 are within 0.5 K at all levels and within ~0.7 K for ERA5 and HWRF. The STDs between HWRF, and DROP and ERA5 at all levels is roughly constant at 1 K. Larger spread exists in HAMSR–ERA5 differences, which shows an STD between 1.5 and 2.3 K.

Specific humidity mean differences from ERA5 (Fig. 6b) are within 0.5 g kg^{-1} for all datasets except the uncorrected HAMSR data. The differences are greatest in the low-levels between 850 and 1000 hPa, with HWRF, DROP and HAMSR all moister than ERA5, indicating a possible dry bias in ERA5 in this layer. For STD, a similar picture as with temperature is found, with greater spread between HAMSR and ERA5 relative to DROP and HWRF.

Figure 6c shows the corresponding relative humidity differences from ERA5. All datasets except the uncorrected HAMSR data show mean agreement with ERA5 to within 5% from the surface to about 500 hPa. The mean differences all become negative above 500 hPa, suggesting a moist bias in ERA5. The STD shows that HAMSR once again has the largest spread, between about 8% and 35%, while HWRF and DROP are similar in magnitude of ~5% to 20%.

The HAMSR data without bias correction (orange profiles in Fig. 6) show considerably larger biases and STD over most of the troposphere compared to the bias-corrected HAMSR data for temperature and specific and relative humidity.

b. *Estimated 3CH error variances and standard deviations*

Here we present the estimated error variances and STD for all four datasets of temperature, specific and relative humidity. We present the error STD results for the three versions of the HAMSR data, but focus on the bias-corrected HAMSR data for all scan angles (the red profiles in Figs. 7 and 8). The estimated error STD for the uncorrected data are larger and are included for comparison in the figures, but are not discussed in detail. The left panels of Fig. 7 shows the mean of the three 3CH error variances and their STDs of temperature, specific and relative humidity. The right panels of Fig. 7 are the estimated mean error STDs, with their STDs indicated by the shading. The estimated error STDs are the square root of the error variances (left panels). Where the estimated error variances are negative, as seen in the ERA5 mean between 500 and 200 hPa and the HWRF and DROP relative humidities above 200 hPa, they are set to zero for the STD estimates. Small negative error variance estimates for a dataset, which are physically impossible, can happen when there are small error correlations between that dataset and one or more of the other datasets (Rieckh and Anthes 2018; Sjoberg et al. 2021). These error correlations may be real, or a result of chance in a small sample size such as in these calculations. Here we discuss only the estimated error STD, because these are easiest to compare with other estimates of errors.

The estimated error STD for temperature for ERA5, HWRF, and DROP (Fig. 7b) exhibit a similar pattern and magnitude, with all three datasets showing a mean error STD of less than 0.8 K throughout most of the troposphere. The ERA5 error estimates are generally slightly smaller than those of HWRF and DROP, averaging around 0.5 K. HAMSR's estimated error STDs are noticeably larger compared to the other three datasets (Fig. 7b), approximately 1.8 K throughout most of the troposphere, but exceeding 2.2 K at 200 hPa. The estimated temperature errors are relatively constant with height, likely because the temperature variability does not vary greatly with height (Figs. 4 and 5).

Similarly, Fig. 7d presents the estimated error STD for specific humidity. As with temperature, ERA5, HWRF, and DROP show similar errors, less than 1.0 g kg^{-1}, with ERA5 exhibiting the smallest errors. HAMSR's specific humidity error estimates, similar to those of temperature, are larger, showing a maximum of about 2.4 g kg^{-1} at 800 hPa. The high value at this level is likely related to the low vertical resolution of the HAMSR data in a region where there is large vertical variability of specific humidity, as indicated by the example profiles shown in Fig. 5. All of the datasets show a decrease of specific humidity errors above about 800 hPa because of the rapid mean decrease of specific humidity above this level (Fig. 4).

The error STD of relative humidity for ERA5, HWRF, and DROP (Fig. 7f) increase slowly from about 5% at the surface to 10% at 400 hPa. HWRF and DROP remain at 10% above this level, while ERA5 shows an increase, reaching a maximum of about 20% at about 250 hPa. In comparison to these three datasets, HAMSR shows over twice the error STD of the other datasets between 1000 and 200 hPa. All four datasets above 150 hPa show large spread and much different error STDs; this could be related to the small amounts of moisture at these levels and consequently the large uncertainties in estimates of relative humidity errors.

The HAMSR data before bias correction (purple and gray profiles in right panels of Figs. 7 and 8) show considerably larger estimated error STD over most of the troposphere compared to the bias-corrected HAMSR data for temperature and specific and relative humidity, indicating the overall higher quality of the bias-corrected HAMSR data. As expected, the uncorrected HAMSR data with scan angles restricted to ≤30° are more accurate than the data that includes all scan angles (Brown et al. 2011).

Figure 8 shows the corresponding mean error variances and STDs of temperature, specific and relative humidity normalized by the ERA5 mean profiles. For temperature (Fig. 8b), the pattern of the STD is similar to that shown in Fig. 7b. ERA5, HWRF, and DROP exhibit errors generally less than 0.2% up to 300 hPa, and up to 0.4% above this level. HAMSR errors are larger, around 0.5% at the surface, increasing to a maximum of about 1% at 300 hPa.

The estimated error STD of normalized specific humidity (Fig. 8d) are very similar for ERA5, DROP, and HWRF in the lower troposphere, increasing from about 5% at the surface to about 15% at 500 hPa. Above 500 hPa, the DROP specific humidity errors are somewhat larger than the ERA5 and HWRF errors. The estimated HAMSR errors in the normalized specific humidity are about twice those of the other datasets.

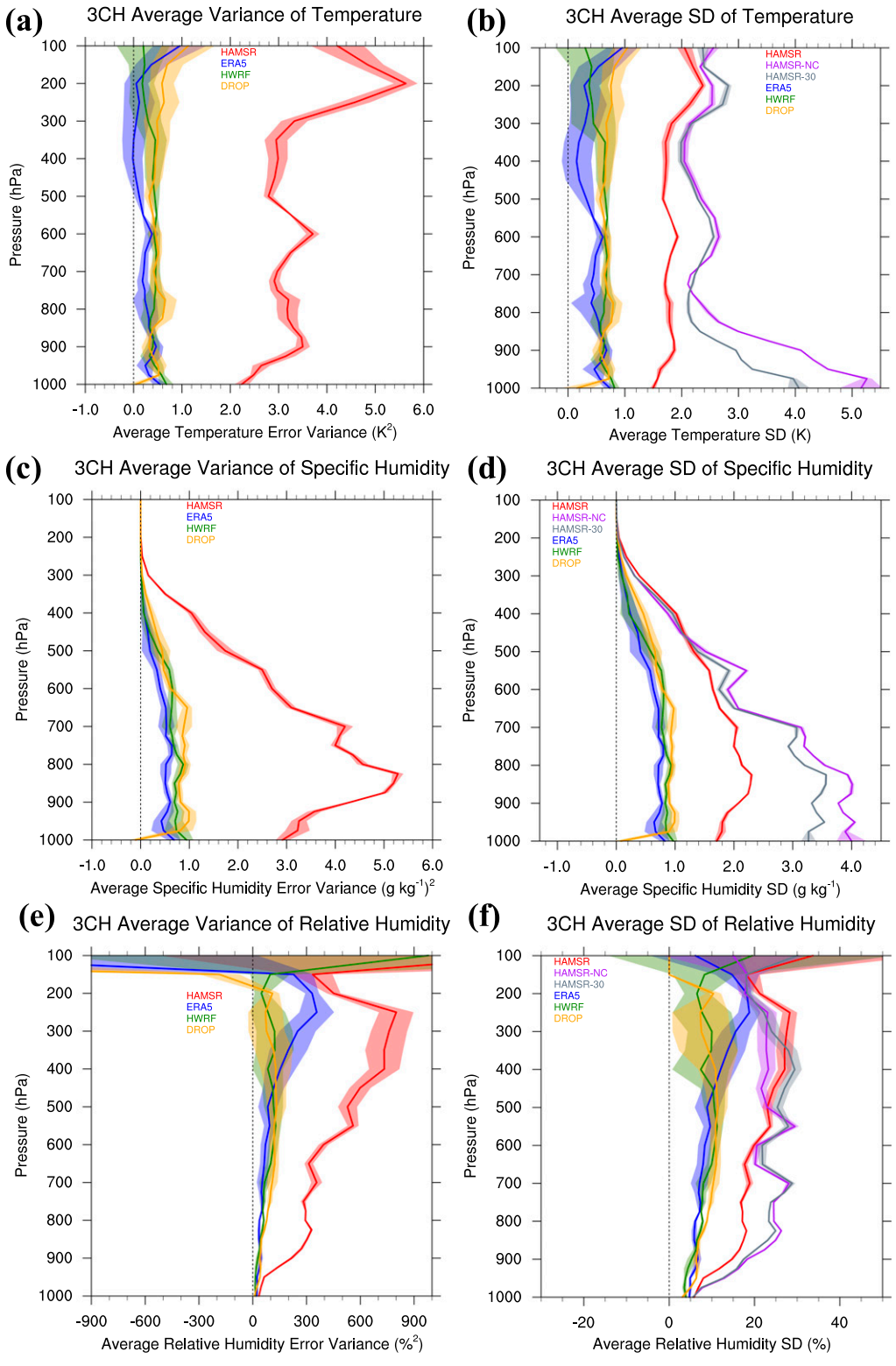

FIG. 7. Estimated error variances using 3CH method for (a) temperature (K^2), (c) specific humidity [$(g\,kg^{-1})^2$], and (e) relative humidity ($\%^2$) for each dataset. The means of the three 3CH estimates are shown by solid lines and the standard deviations of the three estimates [Eq. (9)] are indicated by shading: ERA5 (blue), HAMSR (red), HWRF (green), and DROP (orange). (b),(d),(f) As in (a), (c), and (e), but for the mean estimated 3CH standard deviation and spread for temperature (K), specific humidity ($g\,kg^{-1}$), and relative humidity (%), with HAMSR-NC (all scan angles, not bias corrected; purple) and HAMSR-30 (scan angles less than or equal to 30°, not bias corrected; gray) included.

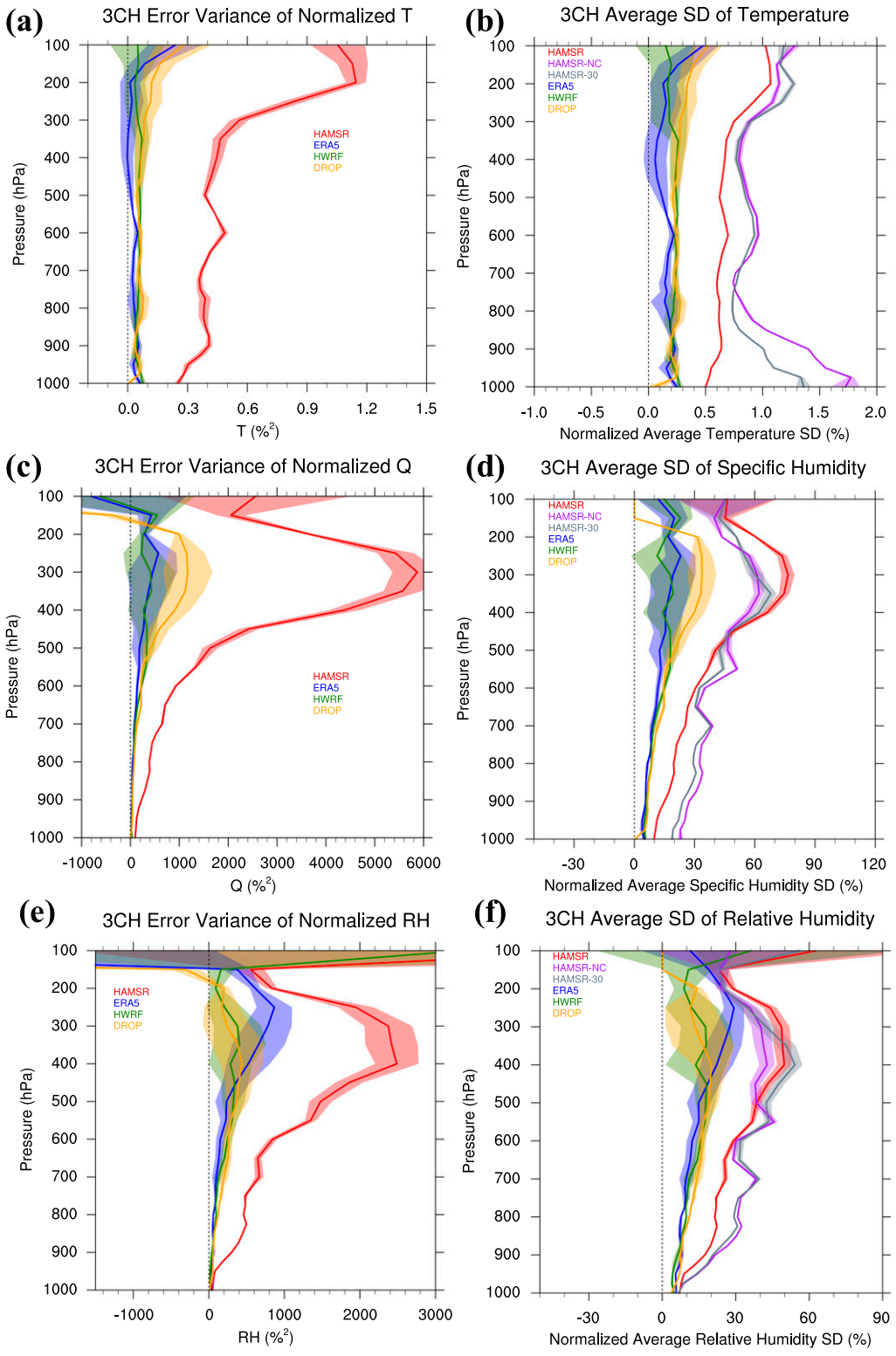

FIG. 8. As in Fig. 7, but for datasets normalized by ERA5 sample means.

For normalized relative humidity (Fig. 8f), the estimated error profiles are similar to those of normalized specific humidity with largest errors in the 200 to 500 hPa levels. HAMSR again shows the largest errors.

c. *Comparison with estimated/known measurement errors*

In this section we compare the results of the estimated 3CH error STD to other estimates of errors of these datasets.

Only two studies, to our knowledge, have evaluated the accuracy of the HAMSR retrievals. Brown et al. (2007) compared the retrievals with nearby dropsondes that were released from the DC-8 aircraft during the NASA African Monsoon Multidisciplinary Analyses campaign (Zipser et al. 2009) in 2006. Comparison with dropsondes on 8 September 2006 (Fig. 1 in Brown et al. 2007) indicated most temperature retrievals were within 2 K of dropsonde measurements, but at some levels differences were as high as 5 K. Absolute humidity from HAMSR was generally within 50% of the dropsonde observations, which is roughly equivalent to a specific humidity difference of 50%.

Brown et al. (2011) compared a retrieved temperature profile to one dropsonde observation in the eye of Hurricane Erin. In this one comparison, differences in the retrieved temperature were generally less than 1 K, but as high as 3–5 K at 2 km and from 12 to 18 km. These prior studies were limited by the number of collocated dropsondes used, but their estimates of temperature errors in the range of 2–5 K and specific humidity of up to 50% are consistent with the 3CH estimates of error STD shown in Fig. 7.

Prior studies estimating the accuracy of the GH dropsondes include Hock and Franklin (1999) and Wick et al. (2018b). Hock and Franklin (1999) estimated the accuracy of the dropsonde temperature observations by examining the measured temperature at the melting level during precipitation. Of 34 comparisons during the Fronts and Atlantic Storm-Track Experiment (Joly et al. 1999), the RMS error was 0.22 K with a bias of 0.09 K, within the estimated accuracy of the instrument (0.2 K in Table 2 of Hock and Franklin, 1999).

Wick et al. (2018b) compared GH dropsondes with dropsondes released from the G-IV during NASA's HS3 campaign (Braun et al. 2016). The dropsondes were deployed over the Gulf of Mexico in 2011 and 2014 during clear to scattered cloud conditions, and were predominately within 2 min and 5 km of each other. Of the 27 collocated dropsondes in 2011 and 15 in 2014, temperature measurements showed mean agreement to within 0.1 K (Fig. 7a in Wick et al. 2018b) and an STD of around 0.4 K. Relative humidity observations showed a mean agreement generally within 5% and an STD of differences largely less than 10% (Fig. 7b in Wick et al. 2018b). These differences are similar to the estimated error in relative humidity of less than 5% shown in Table 3 of Hock and Franklin (1999).

The 3CH results in this paper indicate an error STD of the dropsonde temperature between roughly 0.5 and 0.8 K over all pressure levels (Fig. 7b), which are somewhat higher than the estimates of Wick et al. (2018b) and Hock and Franklin (1999). For relative humidity, our 3CH estimates point to a dropsonde error STD of 5%–10% (Fig. 7f), in good agreement with the 10% STD relative to G-IV dropsondes in Wick et al. (2018b) and about twice the estimated errors of Hock and Franklin (1999). The somewhat higher 3CH error estimates are likely caused by their inclusion of representativeness errors (e.g., Anthes and Rieckh 2018), as the dropsondes are essentially measuring point values while the ERA5 and HWRF models are representing averages over the models' grid volumes, and HAMSR is estimating temperature and humidity over layers of approximately 2 km in depth.

Uncertainty estimates of the ERA5 dataset are provided by the ensemble spread of the ERA5 system. ERA5 includes a reduced resolution (~60 km in the horizontal) 10-member ensemble at a 3-h temporal resolution. The ensemble spread is given as a function of time, latitude, longitude, and the 37 pressure levels (Hersbach et al. 2018). These estimates account for the uncertainty in the observations, model physical parameterizations, and random, but not systematic, errors. Using the provided ensemble spread at each analysis time (every 3 h), we computed a mean and STD of the ensemble spread when closest in time and space to the 533 collocations used in the 3CH estimates.

Figure 9 shows the mean and STD of the ensemble uncertainty for temperature, specific and relative humidity together with our 3CH error STD. The mean ensemble uncertainties of all three variables are somewhat less than the mean 3CH error estimates, but the shapes of the profiles are similar and the STD of the two methods of estimating uncertainty overlap at most levels. The slightly smaller ERA5 ensemble uncertainties may be because the ERA5 ensemble spread does not account for all the uncertainties in the ERA5 dataset (Copernicus Climate Change Service 2017).

We did not find any independent estimates of the error variances or STDs in the temperature, specific humidity, or relative humidity of HWRF forecasts to compare directly to our results in this work.

5. Conclusions

In this study, we used the 3CH method to estimate random error variances and STDs for four datasets in the environment of TCs: ERA5, HWRF analysis, GH dropsondes (DROP), and GH HAMSR retrievals. Prior to this research, no comprehensive study has been done to accurately characterize the error variances and STD of the HAMSR retrievals or to compare them with other datasets. Our study was carried out using data from the 2016 SHOUT season, when the GH sampled TCs Gaston, Hermine, Karl, and Matthew. A total of 533 collocated vertical profiles of the four datasets were found over this time period and used to calculate error estimates.

ERA5, HWRF, and the GH dropsondes all have similar temperature and humidity (specific and relative) errors, with ERA5 having slightly smaller errors then the other datasets. Throughout most of the troposphere, the error STDs of temperature and specific humidity are less than 0.8 K and 1.0 g kg^{-1}, respectively. The estimated error STDs of relative humidity increase from less than 5% near the surface to between 10% and 20% in the upper troposphere.

The close agreement of the dropsonde errors with those of ERA5 and HWRF validates the high accuracy of the GH

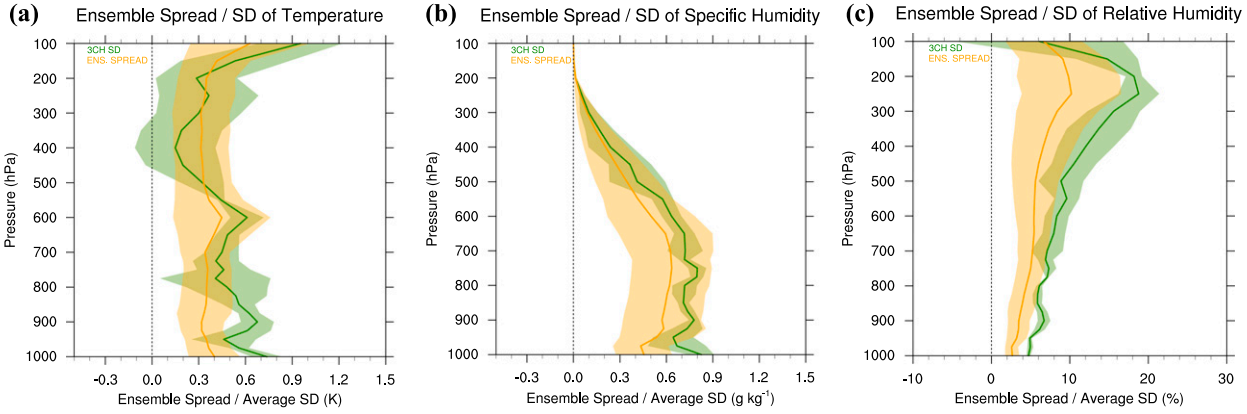

FIG. 9. Estimated mean ensemble spread and standard deviation of ERA5 for points closest in time and space to the 533 collocations used in the 3CH estimates (yellow) and mean estimated 3CH standard deviation and spread (green) for (a) temperature (K), (b) specific humidity (g kg^{-1}), and (c) relative humidity (%).

dropsondes. The estimated error STDs of the dropsondes for temperature (0.5–0.8 K) and humidity (5%–10%) at all pressure levels were found to be similar, but somewhat larger than earlier estimates, likely due to the inclusion of representativeness errors in the 3CH estimates.

HAMSR has larger estimated errors than the other three datasets. The estimated error STDs of temperature and specific humidity in the lower troposphere for the bias-corrected HAMSR data (which include all scan angles) vary between 1.5 and 2.0 K and between 1.5 and 2.5 g kg^{-1}, respectively. The HAMSR error STD of relative humidity increases from approximately 10% in the lower troposphere to 30% in the upper troposphere. The larger errors with respect to the other datasets is likely a result of several aspects. First, HAMSR is remotely sensing the atmosphere, measuring atmospheric brightness temperature. Errors can arise from the retrieval of atmospheric variables, and can be larger when under heavy precipitation. In addition, HAMSR's weighting functions peak only over ~8 levels: 1000, 750, 400, 250, 150, 90, 80, and 40 hPa (Brown et al. 2007). Thus HAMSR could have larger errors in part due to its lower vertical resolution and larger vertical representativeness errors than the other datasets.

We also computed estimated errors associated with the uncorrected HAMSR data and found that the bias-corrected data had smaller errors. We also found that the HAMSR data with scan angles restricted to ≤30° were more accurate than the data that included all scan angles.

We compared our estimated error STDs of ERA5 using the ERA5 ensemble spread data nearest the 533 collocated profiles. Although the mean ensemble uncertainty was less than our estimated error STDs of temperature and specific and relative humidity, possibly because the ERA5 ensemble spread does not account for all the uncertainties in the reanalysis dataset, the STD of the two methods overlapped at most pressure levels.

Acknowledgments. The authors thank Dr. Gary Wick for valuable discussions and collaboration throughout the SHOUT campaign. The authors also acknowledge Dr. Hui Christophersen for valuable input to the manuscript, Dr. Lidia Cucurull for her input and support of this research, and Dr. Sim Aberson for constructive comments and suggestions. This work was funded by the NOAA UAS program in collaboration with the Sensing Hazards with Operational Unmanned Technology project. This research was conducted under the auspices of NOAA's Quantitative Observing System Assessment Program (QOSAP).

Data availability statement. The HAMSR uncorrected retrievals and GH dropsondes were obtained from https://www.esrl.noaa.gov/psd/psd2/coastal/satres/data/static/shout/2016_HRR.html, while the HAMSR bias-corrected data were received directly from Shannon Brown at JPL during the SHOUT campaign. ERA5 was downloaded from Copernicus Climate Change Service (2017), obtained from https://cds.climate.copernicus.eu/cdsapp#!/home. HWRF analyses were obtained from the NCEP EMC archive through the NOAA Environmental Security Computing Center (NESCC) High Performance Storage System.

REFERENCES

Aberson, S. D., K. J. Sellwood, and P. A. Leighton, 2017: Calculating dropwindsonde location and time from TEMP-DROP messages for accurate assimilation and analysis. *J. Atmos. Oceanic Technol.*, **34**, 1673–1678, https://doi.org/10.1175/JTECH-D-17-0023.1.

Anthes, R., and T. Rieckh, 2018: Estimating observation and model error variances using multiple data sets. *Atmos. Meas. Tech.*, **11**, 4239–4260, https://doi.org/10.5194/amt-11-4239-2018.

Braun, S. A., P. A. Newman, and G. M. Heymsfield, 2016: NASA's Hurricane and Severe Storm Sentinel (HS3) investigation. *Bull. Amer. Meteor. Soc.*, **97**, 2085–2102, https://doi.org/10.1175/BAMS-D-15-00186.1.

Brown, S., B. Lambrigsten, A. Tanner, J. Oswald, D. Dawson, and R. Denning, 2007: Observations of tropical cyclones with a 60, 118 and 183 GHz microwave sounder. *IEEE Int. Geoscience and Remote Sensing Symp.*, Barcelona, Spain, IEEE, 3317–3320, https://doi.org/10.1109/IGARSS.2007.4423554.

——, ——, R. F. Denning, T. Gaier, P. Kangaslahti, B. H. Lim, J. M. Tanabe, and A. B. Tanner, 2011: The high-altitude MMIC

sounding radiometer for the Global Hawk unmanned aerial vehicle: Instrument description and performance. *IEEE Trans. Geosci. Remote Sens.*, **49**, 3291–3301, https://doi.org/10.1109/TGRS.2011.2125973.

Chan, P. W., N. G. Wu, C. Z. Zhang, W. J. Deng, and K. K. Hon, 2018: The first complete dropsonde observation of a tropical cyclone over the South China Sea by the Hong Kong Observatory. *Weather*, **73**, 227–234, https://doi.org/10.1002/wea.3095.

Christophersen, H., A. Aksoy, J. Dunion, and K. Sellwood, 2017: The impact of NASA Global Hawk unmanned aircraft dropwindsonde observations on tropical cyclone track, intensity, and structure: Case studies. *Mon. Wea. Rev.*, **145**, 1817–1830, https://doi.org/10.1175/MWR-D-16-0332.1.

——, ——, ——, and S. Aberson, 2018a: Composite impact of Global Hawk unmanned aircraft dropwindsondes on tropical cyclone analyses and forecasts. *Mon. Wea. Rev.*, **146**, 2297–2314, https://doi.org/10.1175/MWR-D-17-0304.1.

——, R. Atlas, A. Aksoy, and J. Dunion, 2018b: Combined use of satellite observations and Global Hawk unmanned aircraft dropwindsondes for improved tropical cyclone analyses and forecasts. *Wea. Forecasting*, **33**, 1021–1031, https://doi.org/10.1175/WAF-D-17-0167.1.

Copernicus Climate Change Service, 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store, accessed 1 August 2019, https://cds.climate.copernicus.eu/cdsapp#!/home.

Desroziers, G., and S. Ivanov, 2001: 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.

Dunion, J. P., G. A. Wick, P. G. Black, and J. Walker, 2018: Sensing hazards with operational unmanned technology: 2015-2016 campaign summary, Final Report. NOAA Tech Memo. OAR-UAS-001, 49 pp., https://doi.org/10.7289/V5/TM-OAR-UAS-001.

Emory, A. E., M. McLinden, M. Schreier, and G. A. Wick, 2015: An introduction to the NASA East Pacific Origins and Characteristics of Hurricanes (EPOCH) field campaign. *Trop. Cyclone Res. Rev.*, **4**, 124–131, http://tcrr.typhoon.org.cn/EN/abstract/abstract105.shtml.

Gilpin, S., T. Rieckh, and R. Anthes, 2018: 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.

Gopalakrishnan, S. G., F. Marks, X. Zhang, J.-W. Bao, K.-S. Yeh, and R. Atlas, 2011: The Experimental HWRF System: A study on the influence of horizontal resolution on the structure and intensity changes in tropical cyclones using an idealized framework. *Mon. Wea. Rev.*, **139**, 1762–1784, https://doi.org/10.1175/2010MWR3535.1.

Gray, J. E., and D. W. Allan, 1974: A method for estimating the frequency stability of an individual oscillator. *28th Annual Symp. on Frequency Control*, Atlantic City, NJ, IEEE, 243–246, https://doi.org/10.1109/FREQ.1974.200027.

Hersbach, H., and Coauthors, 2018: Operational global reanalysis: Progress, future directions and synergies with NWP. ECMWF ERA Rep. Series 27, 65 pp., https://www.ecmwf.int/node/18765.

Hock, T. F., and J. L. Franklin, 1999: The NCAR GPS dropwindsonde. *Bull. Amer. Meteor. Soc.*, **80**, 407–420, https://doi.org/10.1175/1520-0477(1999)080<0407:TNGD>2.0.CO;2.

Joly, A., and Coauthors, 1999: Overview of the field phase of the Fronts and Atlantic Storm-Track Experiment (FASTEX) project. *Quart. J. Roy. Meteor. Soc.*, **125**, 3131–3163, https://doi.org/10.1002/qj.49712556103.

Kren, A. C., L. Cucurull, and H. Wang, 2018: Impact of UAS Global Hawk dropsonde data on tropical and extratropical cyclone forecasts in 2016. *Wea. Forecasting*, **33**, 1121–1141, https://doi.org/10.1175/WAF-D-18-0029.1.

Rieckh, T., and R. Anthes, 2018: Evaluating two methods of estimating error variances using simulated data sets with known errors. *Atmos. Meas. Tech.*, **11**, 4309–4325, https://doi.org/10.5194/amt-11-4309-2018.

Sjoberg, J. P., R. A. Anthes, and T. Rieckh, 2021: The three-cornered hat method for estimating error variances of three or more atmospheric data sets. Part I: Overview and evaluation. *J. Atmos. Oceanic Technol.*, https://doi.org/10.1175/JTECH-D-19-0217.1, in press.

Tallapragada, V., and Coauthors, 2014: Hurricane Weather Research and Forecasting (HWRF) Model: 2014 scientific documentation. NCAR Development Testbed Center Rep., 105 pp., http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.6a_ScientificDoc.pdf.

Wick, G. A., J. P. Dunion, and J. Walker, 2018a: Sensing hazards with operational unmanned technology: Impact study of Global Hawk unmanned aircraft system observations for hurricane forecasting, final report. NOAA Tech Memo. OAR-UAS-002, 94 pp., https://repository.library.noaa.gov/view/noaa/17823.

——, T. F. Hock, P. J. Neiman, H. Vomel, M. L. Black, and J. R. Spackman, 2018b: The NCAR–NOAA Global Hawk dropsonde system. *J. Atmos. Oceanic Technol.*, **35**, 1585–1604, https://doi.org/10.1175/JTECH-D-17-0225.1.

——, and Coauthors, 2020: NOAA's Sensing Hazards with Operational Unmanned Technology (SHOUT) experiment: Observations and forecast impacts. *Bull. Amer. Meteor. Soc.*, **101**, E968–E987, https://doi.org/10.1175/BAMS-D-18-0257.1.

Zipser, E. J., and Coauthors, 2009: The Saharan air layer and the fate of African easterly waves—NASA's AMMA field study of tropical cyclogenesis. *Bull. Amer. Meteor. Soc.*, **90**, 1137–1156, https://doi.org/10.1175/2009BAMS2728.1.