The Three-Cornered Hat Method for Estimating Error Variances of Three or More Atmospheric Datasets. Part I: Overview and Evaluation

JEREMIAH P. SJOBERG,^a RICHARD A. ANTHES,^a AND THERESE RIECKH^a

^a COSMIC Program Office, University Corporation for Atmospheric Research, Boulder, Colorado

(Manuscript received 20 December 2019, in final form 13 October 2020)

ABSTRACT: The three-cornered hat (3CH) method, which was originally developed to assess the random errors of atomic clocks, is a means for estimating the error variances of three different datasets. Here we give an overview of the historical development of the 3CH and select other methods for estimating error variances that use either two or three datasets. We discuss similarities and differences between these methods and the 3CH method. This study assesses the sensitivity of the 3CH method to the factors that limit its accuracy, including sample size, outliers, different magnitudes of errors between the datasets, biases, and unknown error correlations. Using simulated datasets for which the errors and their correlations among the datasets are known, this analysis shows the conditions under which the 3CH method provides the most and least accurate estimates. The effect of representativeness errors caused by differences in vertical resolution of datasets is investigated. These representativeness errors are generally small relative to the magnitude of the random errors in the datasets, and the impact of this source of errors can be reduced by appropriate filtering.

KEYWORDS: Atmosphere; Occultation; Error analysis; Model errors

1. Introduction

Observations are essential for understanding the different components of the Earth system. They are the basis for many applications that support warning systems of hazards to life and property, such as numerical weather prediction (NWP) models. For these applications, it is important to understand the errors associated with in situ and remotely sensed observational systems. Many diverse observations are used to initialize forecast models or produce long-term reanalyses of the Earth system through nearly continuous assimilation of observations into the models (Simmons et al. 2016). In modern data assimilation, each type of observation is weighted inversely relative to the magnitude of its expected random errors (uncertainties), expressed statistically as an error variance (e.g., Daley 1993; Desroziers and Ivanov 2001; Weston et al. 2014; Bormann et al. 2016). These errors include all sources of errors: instrument, representativeness, forward modeling, and processing errors including quality control. The error statistics may vary strongly with geographic location, latitude, season, and synoptic situation. Comparison of different observational techniques and models and how they improve over time requires estimating the errors of their respective datasets.

A powerful technique for simultaneously estimating random error statistics of three datasets, called the three-cornered hat (3CH) method, is the simplest case ($N = 3$) of the more general N -cornered hat method, which uses N datasets. This method was developed by Grubbs (1948) to estimate the errors of three different instruments. While widely used in the physics community, the 3CH method has not been used to estimate the

errors of atmospheric datasets until recently (Anthes and Rieckh 2018; Rieckh and Anthes 2018).

A related method involving only two datasets, termed the two-cornered hat (2CH) method by Rieckh and Anthes (2018), was used by Braun et al. (2001) to estimate the random errors associated with ground-based global positioning system (GPS) receivers and water vapor radiometers. An important method similar to the 3CH method, known as triple collocation (TC), was developed independently of the 3CH method (Stoffelen 1998), and has been widely used to estimate the errors of ocean, land, and biological datasets. While superficially similar, the 2CH, 3CH, and TC methods actually contain differences in assumptions, and the origins and relationships among them have, to our knowledge, never been systematically analyzed. This is also true of the properties and limitations of the 3CH method.

In this paper, we carry out such an analysis for the 3CH method, beginning with a discussion of its history. We then show a brief derivation of the 3CH error variance equations, and compare these with commonly used metrics of error estimates using two datasets (apparent error and standard deviation of differences) and with the 2CH and TC methods. We follow this with a detailed exploration of the properties of the 3CH method. These include the factors that potentially limit the accuracy of the 3CH method: correlations of errors between two or more of the datasets, different magnitudes of random errors, outliers, biases, and sample size. We conclude with a discussion of how the definition of truth for each set of observations affects the interpretation of the 3CH results, extending the work of O'Carroll et al. (2008), and how the definition relates to representativeness errors between datasets with different vertical scales.

2. History of the 3CH method

We traced the history of the 3CH method back to Grubbs (1948), who discusses various methods of comparing two,

Denotes content that is immediately available upon publication as open access.

Corresponding author: Jeremiah P. Sjöberg, sjoberg@ucar.edu

three, and N independent datasets. Grubbs assumes at the beginning that there is no correlation of errors in the datasets, and derives the equations for what we now call the 3CH and 2CH methods, although he does not use these terms.

Barnes (1966) presented the three linearly independent equations relating the error variances of three atomic oscillators [his Eq. (18)] that are the basis for the 3CH equations, even though he did not use the term 3CH nor did he write down the 3CH equations that are derived from his Eq. (18). Levine (1999) in her review of time and frequency metrology mentions “the three-cornered hat” method and references Barnes (1966).

The next 3CH paper was that of Gray and Allan (1974), who apparently derived the 3CH equations independently of Grubbs (1948) and used it for estimating the random errors of three or more oscillators. Gray and Allan (1974) called the method the “triangulation method,” and assumed no correlation of errors among the oscillators being tested. They indicated that the error variance of one oscillator may be evaluated even if the errors associated with two other oscillators are larger. They also noted that if the errors of an oscillator were small enough, the 3CH equations could occasionally result in small negative numbers, which is physically impossible. Riley (2008) summarized the 3CH method for estimating the errors of three or more oscillators and states in a footnote: “the term ‘Three-cornered hat’ was coined by J. E. Gray of NIST.”

By the mid-1980s the 3CH method had become fairly widely used in the physics community. Greenhall (1987) called it “the classical 3-cornered hat” method and proposed a solution using a statistical model based on the method of maximum likelihood. Premoli and Tavella (1993) advanced “the popular three-cornered hat method” by addressing the issue of occasional negative error variance estimates by accounting for the possible correlation of errors in a statistical minimization technique. This generalization was further developed by Tavella and Premoli (1994) and Galindo et al. (2001). Ekstrom and Koppang (2006) state, “a common technique is to use phase or frequency measurements between three (or more) oscillators in a procedure commonly referred to in the timing community as a three- (or n -) cornered hat.” Griggs et al. (2014) used the 3CH method to estimate the stability of GNSS clocks, but used the Gray and Allan name for it, “triangulation method.” In a later paper, Griggs et al. (2015) used both terms. In both papers they say that the accuracy of the 3CH method is limited if one of the datasets has errors much larger than the other two.

Anthes and Rieckh (2018) and were the first to use the 3CH method for atmospheric datasets, comparing radiosonde and radio occultation (RO) soundings, ERA-Interim (Dee et al. 2011), and the National Centers for Environmental Prediction Global Forecast System model. Rieckh and Anthes (2018) compared the 3CH and 2CH method using simulated datasets for which the truth was known and concluded that the 2CH method was much more sensitive to random errors and sample size than the 3CH method. For this reason, our work here focuses on the properties of the 3CH method and the factors that limit its accuracy.

Stoffelen (1998) independently introduced a method for estimating the errors associated with three datasets, using one

of the datasets to calibrate the other two. He called this method the “triple co-location” (TC) method, and used it to estimate the error variances of sea surface winds. His paper was followed by many other studies. For example, McColl et al. (2014) state:

TC has been used widely in oceanography to estimate errors in measurements of sea surface temperature (Gentemann 2014; O’Carroll et al. 2008), wind speed and stress (Portabella and Stoffelen 2009; Stoffelen 1998; Vogelzang et al. 2011), and wave height (Caires and Sterl 2003; Janssen et al. 2007). It has also been used in hydrology to estimate errors in measurements of precipitation (Roebeling et al. 2012), fraction of absorbed photosynthetically active radiation (D’Odorico et al. 2014), leaf area index (Fang et al. 2012), and, particularly, soil moisture (Anderson et al. 2012; Dorigo et al. 2010; Draper et al. 2013; Hain et al. 2011; Miralles et al. 2010; Parinussa et al. 2011; Scipal et al. 2008; Su et al. 2014). It has been applied in data assimilation (Crow and van den Berg 2010) and can also be used to optimally rescale two measurement systems to a third reference system (Stoffelen 1998; Yilmaz and Crow 2013).

The above studies use one of two different and identical forms of the TC equations, which Gruber et al. (2016) term the difference and covariance notations.

O’Carroll et al. (2008) used the 3CH equations without calibration or scaling of the datasets—probably deriving them independently of Gray and Allan (1974) and following papers—to estimate the errors between three types of sea surface temperature observations. This important paper clearly describes the 3CH method and its limitations, and discusses representativeness errors and the meaning of “truth.” However, it does not mention the 3CH method nor reference any of the papers using the 3CH method. Instead they use the term “three-way collocation” and reference Stoffelen (1998) and Blackmore et al. (2007). The Blackmore et al. (2007) reference is a Met Office technical report that used the 3CH method, referencing a 2006 draft of the O’Carroll et al. (2008) paper.

None of the TC papers that we could find in the geosciences literature uses the term 3CH, nor do they reference the earlier papers in the physics community that use the 3CH equations. And none of the 3CH papers published after the Stoffelen (1998) paper reference the TC method. In summary, the 3CH, 2CH, and TC methods were apparently developed independently and in parallel.

3. Summary of the 3CH method

a. The 3CH error variance equations

Consider the set of observations X comprised of data points at given locations and times. We can write this set of observations as

$$X = T + \beta_X + \varepsilon_X, \quad (1)$$

where T is the set of true, unknown values in X , or “truth”; β_X is the mean bias of X relative to T ; and ε_X is the set of random errors associated with X . The error variance of X is defined as

$$\text{Var}_{\text{err}}[X] := E[(X - T - \beta_X)^2] = E[\varepsilon_X^2], \quad (2)$$

where $E[\cdot]$ denotes the sample mean. For three datasets X , Y , and Z for which the form of Eq. (1) holds, it can be shown (appendix A) that the exact equations for the error variances are

$$\begin{aligned} \text{Var}_{\text{err}}[X] = & \frac{1}{2} \{E[(X - Y)^2] + E[(X - Z)^2] - E[(Y - Z)^2]\} \\ & - \frac{1}{2}(b_{XY}^2 + b_{XZ}^2 - b_{YZ}^2) \\ & + E[\varepsilon_X \varepsilon_Y] + E[\varepsilon_X \varepsilon_Z] - E[\varepsilon_Y \varepsilon_Z], \end{aligned} \quad (3a)$$

$$\begin{aligned} \text{Var}_{\text{err}}[Y] = & \frac{1}{2} \{E[(X - Y)^2] + E[(Y - Z)^2] - E[(X - Z)^2]\} \\ & - \frac{1}{2}(b_{XY}^2 + b_{YZ}^2 - b_{XZ}^2) \\ & + E[\varepsilon_X \varepsilon_Y] + E[\varepsilon_Y \varepsilon_Z] - E[\varepsilon_X \varepsilon_Z], \end{aligned} \quad (3b)$$

$$\begin{aligned} \text{Var}_{\text{err}}[Z] = & \frac{1}{2} \{E[(X - Z)^2] + E[(Y - Z)^2] - E[(X - Y)^2]\} \\ & - \frac{1}{2}(b_{XZ}^2 + b_{YZ}^2 - b_{XY}^2) \\ & + E[\varepsilon_X \varepsilon_Z] + E[\varepsilon_Y \varepsilon_Z] - E[\varepsilon_X \varepsilon_Y], \end{aligned} \quad (3c)$$

where the b terms here indicate mean biases between two datasets (e.g., $b_{XY} = E[X - Y] = \beta_X - \beta_Y$). Equations (3)—the 3CH error variance equations—are equivalent to Eq. (A8) in O'Carroll et al. (2008). The standard deviation (SD) of the errors is the square root of the error variance.

With three datasets, we can compute all the terms in Eqs. (3) except the error covariance terms, and so they are assumed to be zero. Under ideal conditions with error independence—i.e., all error covariances equal to zero—the above equations are exact despite the neglect of these terms. If two or more of the datasets have correlated errors, some of the neglected error covariance terms are nonzero and the estimated error variances will differ from the true error variances. Thus, as discussed below, correlation of errors is one of the factors that limit the accuracy of the 3CH method.

The derivation for the 3CH relations in Eqs. (3) generalizes to N datasets (appendix A). That is, the 3CH is a special case of the N -cornered hat where $N = 3$. For more than three datasets ($N > 3$), there are $(N - 1)(N - 2)/2$ sets of unique 3CH error variance relations for each dataset, and therefore we can obtain $(N - 1)(N - 2)/2$ estimates of the error variance for each dataset (appendix A). If all error covariances are equal to zero, the different estimates of the error variances will be identical. With nonzero error covariances, the estimates will be different and the spread of the estimates is an indicator of the magnitude of the error correlations and thus the uncertainty of the estimates.

b. Relationship to other methods of estimating errors

Many studies estimate the errors of one dataset by comparing it with another dataset, usually one that is considered quite accurate or one in which we have an independent estimate of the error variance. Examples include carefully calibrated radiosondes, or model analysis, background, or reanalysis.

Kuo et al. (2004) and Chen et al. (2011) estimate the apparent error of dataset X as

$$X_{\text{AE}} = X - Y, \quad (4)$$

where Y is a trusted dataset, e.g., a model analysis or short-term forecast. If Y is a model's background, the apparent error is the observation minus background or $O - B$ (Desroziers et al. 2005) that is used in NWP centers to monitor the quality of observations. For X and Y for which Eq. (1) holds, the variance of the apparent error is

$$\text{Var}[X_{\text{AE}}] = E[(X - Y)^2] - b_{XY}^2 = \text{RMS}[X - Y]^2 - b_{XY}^2. \quad (5)$$

Thus, the standard deviation of the differences will always be less than or equal to the root-mean-square (RMS) difference.

Equation (5) is equivalent to

$$\text{Var}[X_{\text{AE}}] = \text{Var}_{\text{err}}[X] + \text{Var}_{\text{err}}[Y] - 2E[\varepsilon_X \varepsilon_Y]. \quad (6)$$

We note from Eq. (6) that the variance of the apparent error for X (or differences between X and another dataset Y) will always be greater than the 3CH estimate of the $\text{Var}_{\text{err}}[X]$ if there is a negligible covariance of errors between X and Y .

The 2CH also involves only two datasets. From Rieckh and Anthes (2018), the 2CH error variance of X analogous to Eq. (3a) is

$$\begin{aligned} \text{Var}_{\text{err}}[X] = & E[X^2] - E[XY] \\ & - \frac{1}{2}(b_{XY}^2 + E[X]^2 - E[Y]^2) \\ & + E[\varepsilon_X \varepsilon_Y] - E[T(\varepsilon_X - \varepsilon_Y)]. \end{aligned} \quad (7)$$

In applications of both the 3CH and 2CH methods, we must neglect the error covariance terms. For the 2CH relation in Eq. (7), the unknown mean of the product of truth T and the difference of errors in the datasets must also be ignored. This makes an implicit assumption about the relationship between truth and the error terms ε . This assumption that the mean of the product of truth and error terms is zero, termed error orthogonality, is not necessary for the 3CH method.

When these terms are neglected, and assuming no biases and equal means for each dataset, the estimated error variance equations for dataset X from the 3CH and 2CH methods, Eqs. (3a) and (7), respectively, may be rearranged as

$$\text{3CH: } \text{Var}_{\text{err}}[X]_{\text{est}} = E[X^2] - E[XY] - E[Z(X - Y)], \quad (8a)$$

$$\text{2CH: } \text{Var}_{\text{err}}[X]_{\text{est}} = E[X^2] - E[XY]. \quad (8b)$$

Equations (8) are identical except for the extra term $E[Z(X - Y)]$ involving the third dataset Z in the 3CH method. Including this term, and the impact of the neglected terms, makes a large difference in the results from the two methods, as shown by Rieckh and Anthes (2018):

- Random errors have a much larger effect on the 2CH method, and so the 2CH method is much more sensitive to sample size.
- If the between-dataset bias terms b in both methods are neglected, as in Rieckh and Anthes (2018), the 2CH method is more sensitive to biases between the two datasets.

Unlike the above methods, the TC method uses three datasets, though it casts the error model given by Eq. (1) in the alternate form

$$X = \alpha_X T + \beta_X + \varepsilon_X.$$

Note that $\alpha_X \neq 1$ implies that dataset X contains multiplicative biases, i.e., it is uncalibrated. We further note that the TC method also assumes error orthogonality. In general, these differences from the 3CH error model and assumptions make the TC method distinct. Both methods require the assumption of error independence in application, however.

Considering the above error model, together with the TC method assumption of error orthogonality, it can be shown (appendix B) that the corresponding 3CH error variance equation for X with multiplicative biases is

$$\begin{aligned} \text{Var}_{\text{err}}[X] = & \frac{1}{2} \{E[(X - Y)^2] + E[(X - Z)^2] - E[(Y - Z)^2]\} \\ & - \frac{1}{2} (b_{XY}^2 + b_{XZ}^2 - b_{YZ}^2) \\ & + E[\varepsilon_X \varepsilon_Y] + E[\varepsilon_X \varepsilon_Z] - E[\varepsilon_Y \varepsilon_Z] \\ & - \alpha_X \left(\alpha_X - \alpha_Y - \alpha_Z + \frac{\alpha_Y \alpha_Z}{\alpha_X} \right) E[T'^2], \end{aligned}$$

where $T' = T - E[T]$ is the anomaly of truth from its mean.

The above multiplicative bias form of the 3CH contains a term that includes both the unknown parameters α and unknown variance of truth. Through rescaling by one of the datasets, the TC method mitigates the effect of these terms (e.g., Gruber et al. 2016). For three datasets that are similarly (un)calibrated to each other and for which the α terms are approximately equal to one, the effect of the multiplicative bias terms will be negligible. An example of this is given in Table 6 of Vogelzang et al. (2011), who show that the ratios of α terms in a particular set of scatterometer, model, and buoy data are all close to one. Under such conditions, the last term in the above error variance relation is approximately zero and we recapture the form of the 3CH relation for X with no multiplicative biases as given in Eq. (3a). In general, however, uncalibrated data may not have similar values of α , and the TC and 3CH methods will be distinct. A review of the literature indicates that the multiplicative bias coefficients are most variable for soil moisture datasets [e.g., Loew and Schlenz (2011), in which the α terms range from 0.24 to 0.95]. In such cases the 3CH method error estimates may have large errors and the TC method should be used instead.

In sensitivity studies comparing the 3CH and TC methods using synthetic datasets with various values of α , we find that the multiplicative biases have small effects on the 3CH methods when α varies between 0.95 and 1.05. For larger α , we find greater differences and the TC method gives superior results. We also compared the 3CH and TC methods for three atmospheric sets: Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) radio occultation, European Centre for Medium-Range Weather Forecasts reanalysis fifth generation (ERA5), and Modern-Era Retrospective Analysis for Research and Applications, second generation (MERRA-2). We found that the 3CH and TC results were almost identical,

indicating that the effect of multiplicative biases in these atmospheric datasets were negligible.

In addition to the particular case above, a similarity between the methods is shown in appendix A of Gruber et al. (2016). After rescaling, the TC error model takes the form of the 3CH error model, though with a constant factor on truth for all three datasets [their Eq. (A.5)]. The final step of the TC method, then, is to calculate error variance in the same way as the 3CH method. Note that this does not imply equivalence because the TC method requires rescaling that itself requires the assumptions of error orthogonality and error independence.

To summarize, the three methods are distinct: 2CH uses only two datasets while the 3CH and TC methods use three datasets in their error estimation. The TC method requires the assumption of error orthogonality, but can handle uncalibrated data. The 3CH method requires the assumption that data are not wildly uncalibrated, but does not require the assumption of error orthogonality.

4. Factors that limit the accuracy of the 3CH method

There are a number of factors that potentially limit the accuracy of the 3CH method, and therefore the reliability of the error variance and standard deviation estimates:

- (i) sample size,
- (ii) outliers in one or more of the datasets,
- (iii) magnitude of the random errors in the datasets and a different magnitude of errors in one of the datasets compared to the other two,
- (iv) biases from truth in the datasets, and
- (v) unknown error correlations.

The five factors above are not independent. For example, the sensitivity to sample size may depend on the other factors, such as the magnitude of the random errors. In the derivation of the 3CH method, many terms (e.g., the product of means and random errors) drop out for an infinite sample size. But for a finite sample size, these neglected terms will not be exactly zero. For finite sample sizes, small error correlations may arise by chance, even if the three datasets are completely independent.

We additionally note that all of the five factors that may affect the accuracy and interpretation of the 3CH method also may affect the accuracy and interpretation of the 2CH method and other methods of comparing two or more datasets as discussed in section 3b. Indeed, similar sensitivity analyses have been performed for the TC method (e.g., Vogelzang and Stoffelen 2012; Zwieback et al. 2012; Yilmaz and Crow 2013).

To independently demonstrate the relative impact of each of these factors, we simulate three datasets with different random errors or biases that we specify.

We begin by choosing an arbitrary, but realistic atmospheric dataset of profiles as truth, to which errors are added to create three different datasets. For this truth dataset, we used profiles from the ERA-Interim dataset from a region bounded by 10°S–10°N latitude and 40°–50°E longitude for the month of January 2007. Using 6-hourly data, this gives us 52 080 profiles. The data are given on 37 pressure levels (Berrisford et al. 2011, Table 2).

FIG. 1. Mean (black profile) and ± 1 standard deviation (shading) of the 52 080 truth profiles for (left) refractivity (N-units), (center) specific humidity (g kg⁻¹), and (right) temperature (K).

The vertical grid spacing is 25 hPa from 1000 to 750 hPa and from 225 to 100 hPa, 50 hPa from 700 to 250 hPa, and varying spacing above 100 hPa.

All data are normalized using the sample average of truth for refractivity, specific humidity, and temperature. The means and ± 1 standard deviation of the profiles are shown in Fig. 1. We emphasize results for specific humidity because of its large variability in the atmosphere and the large variation in errors expected using different observing systems and models (Dee and da Silva 2003; Schröder et al. 2018). We include refractivity N because it is a widely used observable from RO and easily computed from NWP models according to Smith and Weintraub (1953):

$$N = 77.6 \frac{p}{T} + 3.73 \times 10^5 \frac{e}{T^2}, \quad (9)$$

where p is pressure (hPa), T here is temperature (K), and e is water vapor pressure (hPa).

Random errors are normally distributed around truth with a standard deviation given by the respective error model. The model error standard deviations (SD_{err}) are given as follows:

- Refractivity: Specified with pressure by an empirical percentage of the mean true profile N value based on the general shape of previous estimates of RO error profiles (e.g., Kuo et al. 2004).
- Specific humidity: Increase with decreasing pressure as a percentage of the mean true profile q value according to $SD_{\text{err}} = [0.021(1000 - p) + 5]q$, where p is the pressure in hPa.
- Temperature: Constant with pressure at 0.6 K.

FIG. 2. Exact error standard deviations (top) in physical units and (bottom) in normalized units specified in the error models for (left) refractivity, (center) specific humidity, and (right) temperature.

FIG. 3. (left) Exact (solid) and 3CH estimated (dashed) normalized specific humidity error standard deviation (%) for the control experiment with three synthetic datasets. (right) Vertical profiles of normalized error covariance terms (%) from this experiment. Error correlations between the datasets are zero. The nonzero values, which are neglected in the 3CH method, arise purely from chance in a sample of finite size and increase in magnitude with decreasing pressure according to our error model.

The error model standard deviation values, in physical and normalized units for refractivity, specific humidity, and temperature are shown in Fig. 2.

The actual random error for each point in each dataset is computed as

$$R_i = A_i \times \text{SD}_{\text{err}} \times \text{RAND}, \quad (10)$$

where A_i is a factor that controls the overall magnitude (default $A_i = 1$) and RAND is a random number from a normal distribution with a mean of zero and a standard deviation of one. The i th dataset here refers to one of X , Y , or Z .

We may correlate any two of the datasets i and j by expressing the error of dataset i as a linear combination of R_i and R_j :

$$\varepsilon_i = \frac{\sum_j a_{ij} R_j}{\sum_j a_{ij}}. \quad (11)$$

In the case of three datasets X , Y , and Z , Eq. (11) can be expanded as

$$\varepsilon_X = \frac{[R_X + a_{XY}R_Y + a_{XZ}R_Z]}{[1 + a_{XY} + a_{XZ}]}, \quad (12a)$$

$$\varepsilon_Y = \frac{[a_{YX}R_X + R_Y + a_{YZ}R_Z]}{[a_{YX} + 1 + a_{YZ}]}, \quad (12b)$$

$$\varepsilon_Z = \frac{[a_{ZX}R_X + a_{ZY}R_Y + R_Z]}{[a_{ZX} + a_{ZY} + 1]}. \quad (12c)$$

FIG. 4. (left) Exact (solid) and 3CH estimated (dashed) normalized refractivity error standard deviation (%) for the control experiment with three synthetic datasets. (right) Vertical profiles of normalized error covariance terms (%) from this experiment. Error correlations between the datasets are zero. The nonzero values, which are neglected in the 3CH method, arise purely from chance in a sample of finite size and decrease in magnitude with decreasing pressure according to our error model.

FIG. 5. (left) The effect of a smaller sample size of 500 profiles on exact (solid) and 3CH estimated (dashed) normalized specific humidity error standard deviation (%). (right) Vertical profiles of normalized error covariance terms ($\%^2$) from this experiment. Due to a smaller sample size, coincidental nonzero error correlations increase in magnitude and create larger errors in the estimated error standard deviations.

With our error models, there is no simple relationship between the correlation coefficient r_{ij} for the errors of the i th and j th datasets the way there was in the simple error model of Rieckh and Anthes (2018). Also, note that $a_{ij} = 0$ does not imply necessarily that there is zero correlation between the errors of datasets i and j . For example, a_{ij} could be zero but a_{ji} could be nonzero. Or a correlation between ε_i and ε_j could occur through their mutual correlation with a third dataset. Thus, the error correlations of the simulated datasets must be computed from the datasets themselves after the a_{ij} are specified.

For the control experiment, we select a random sample of 5000 profiles and then assign random, uncorrelated errors with $A_i = 1$ in Eq. (10) and $a_{ij} = 0$ for $i \neq j$ in Eq. (11). Figures 3 and 4 show the vertical profiles of the exact and estimated error variances for specific humidity and refractivity, respectively. Correlations of the errors among the three datasets are nonzero only because of the finite sample size. The error covariances are therefore nonzero, and the estimated error variances (dashed profiles) depart slightly from the exact error variances (solid profiles). This departure is larger at lower pressures for specific humidity (Fig. 3) since our error model specifies that the magnitudes of the random errors of specific humidity increase with decreasing pressure. For refractivity (Fig. 4), the specified errors decrease with decreasing pressure and thus so do the error covariances.

a. Effect of sample size

For smaller sample sizes, the correlation of errors that arise by chance become larger and so the departures of the estimated error variances from the exact values become larger. Figure 5 shows the results from an experiment identical to the control except the randomly selected sample size is reduced to 500. The larger error covariance terms create larger inaccuracies in the estimated error standard deviations. Increasing the randomly selected sample size to 50 000 meanwhile reduces the magnitude of these spurious covariance terms (not shown).

In general, we conclude that for errors typical of specific humidity, the sample size should be at least 500 for meaningful results and preferably over 5000. Of course, this number varies with datasets, but indicates that care should be taken when the sample size is 500 or less. The sensitivity of the error estimates to sample size can be tested by reducing the full sample to a smaller number, say 50%, of the total number of triplets available.

b. Effect of outliers

To test the effect of outliers in one dataset on all error variance estimates, we repeated the control except we increased

FIG. 6. The effect of large outliers in one of the datasets on exact (solid) and 3CH estimated (dashed) normalized specific humidity error standard deviations (%). The error standard deviations for the 10 largest error values of X are increased by 20 times the original at the 225 hPa level. The estimated and exact error standard deviations of X agree very closely—solid and dashed blue lines are indistinguishable.

FIG. 7. (left) The effect of the magnitude of random errors on exact (solid) and 3CH estimated (dashed) normalized specific humidity error standard deviation (%). (right) Vertical profiles of normalized error covariance terms ($\%^2$) from this experiment. The magnitude of random errors for Z are increased (top) by a factor of 2 ($A = 2$) and (bottom) by a factor of 10 ($A = 10$). The exact and estimated values of error standard deviation for Z in the bottom-left panel are off the scale and not shown.

the error standard deviations for the 10 largest error values of dataset X by a factor of 20 at the 225 hPa level. The results with these 10 outliers are shown in Fig. 6. The 3CH method accurately reveals this increase for X for both the estimate and exact error standard deviations—the solid and dashed blue profiles are indistinguishable in Fig. 6—while the estimates for Y and Z are not affected.

c. Effect of magnitudes of random errors and different orders of magnitude

We expect the accuracy of the 3CH method to decrease as the magnitude of the random errors increases, because the small correlation of errors in finite datasets results in increasing error covariance terms as the magnitude of errors increase. We find that for a modest increase in the magnitude of the random errors for one of the datasets—e.g., from the Control ($A = 1$; Fig. 3) to errors doubled ($A = 2$; top panels of Fig. 7)—the 3CH estimates are nearly as accurate for all datasets. However, in line with earlier studies (e.g., Griggs et al. 2014, 2015), when the errors of one dataset are one or more orders of magnitude larger than the others, the estimated error profiles become

noticeably noisier (bottom panels of Fig. 7). As sample size increases, the spurious error correlations decrease and the 3CH estimates become more accurate (not shown). These experiments indicate that modest (e.g., factors of 2 to 3) differences in magnitude of random errors do not significantly affect the 3CH results with finite sampling.

d. Effect of biases among the datasets

To confirm that biases among the datasets do not affect the 3CH error variance estimates, we carried out several experiments in which mean biases β of up to 20% were added to various combinations of the datasets. We obtained exactly the same error variance estimates for all combinations of biases we studied.

We then looked at the effect of various combinations of biases among the datasets if the bias terms b in Eqs. (3) were neglected in the error variance estimates (as in Anthes and Rieckh 2018). According to Eqs. (3), omission of a bias term b will result in an error of order b^2 . With biases among two of the datasets, there may be partial cancellation of the bias terms. The results from not accounting for a 1% bias added to dataset

FIG. 8. (left) The effect of neglected biases on exact (solid) and 3CH estimated (dashed) normalized specific humidity error standard deviation (%). (right) Vertical profiles of normalized error covariance terms ($\%^2$) from this experiment. A 1% bias is added to dataset Z and the bias terms b are omitted in Eqs. (3).

Z for refractivity are shown in Fig. 8. Omission of the bias terms does not affect the error variance estimates for X and Y because the bias terms cancel in Eqs. (3a) and (3b). The bias terms for Z in Eq. (3c) equal $-b^2$ and their omission results in an overestimate of the error variance by b^2 (an overestimate of the error standard deviation by b).

e. Effect of unknown error correlations

As has been recognized since the first papers on the 3CH method, the effect of unknown error correlations among two or more of the datasets is potentially the most serious drawback to the 3CH method. This is true in part because even small correlations can have a noticeable effect (e.g., Fig. 7) and because it is difficult to determine the degree of correlation independent of the 3CH method. Rieckh and Anthes (2018) investigated the effect of error correlations between two of the datasets, and showed that neglect of error covariances between two of the datasets resulted in underestimates of the errors of the two correlated datasets and overestimates of the errors of the third dataset. For a correlation coefficient of 0.2, the ratio of the estimated to true error standard deviations was approximately 0.9 for the two correlated datasets and about 1.1 for the uncorrelated dataset (Rieckh and Anthes 2018, their Fig. 4b).

Figure 9 shows the results of adding a 0.1, 0.2, or 0.4 correlation between errors in Y and Z. As discussed above, the estimated error standard deviations of the two correlated datasets Y and Z become too small and the estimates of X become too large as the correlation increases.

If the errors of one dataset are correlated with the errors of both of the other datasets, the effects of the error correlations become more complicated: all error covariance terms in Eqs. (3)—both positive and negative—are nonzero. With real datasets, some of the error covariances will thus tend to add while others will tend to cancel. An example is shown in Fig. 10 in which the errors of Z are correlated with those of both X and Y at $r = 0.1$ (top), and with the errors of X at $r = 0.1$ and Y at $r = 0.2$ (bottom row). In the first case (top row), the nonzero

error covariance terms between X and Z and those between Y and Z cancel in the estimates of error standard deviations for X and Y, and are nonzero and additive in the estimate of the error standard deviation of Z. In the second case (bottom row), only partial cancellation occurs and all error estimates are affected.

Gray and Allan (1974) noted that error correlations among the three datasets may lead to physically impossible negative error variance estimates. We simulated the conditions that lead to negative error variances. A negative estimated error variance is favored when the sum of the neglected error covariance terms for a dataset is positive and larger in magnitude than its actual error variance. This may occur from real error correlations in the datasets or correlations that occur by chance due to the finite data size, and from datasets with different magnitudes of errors. An example is shown in Fig. 11, where the errors of Y and Z are highly correlated and the error variance of Z is greater than that of both X and Y. The neglected error covariance terms in each of Eqs. (3) is $E[\varepsilon_Y \varepsilon_Z]$, causing an overestimate in the estimate of X and underestimates of Y and Z, all by the same magnitude. Without the different magnitudes of errors between datasets, the error variance of X remains positive (not shown).

In summary, error correlations of as low as 0.1 can have a noticeable effect on the estimated error standard deviations compared to the true error SD, but the errors in the estimated error SD will be less than 10%. With higher error correlations, say 0.4, the estimated error standard deviations can be up to 40% under or overestimated. For datasets that have modestly different error variance magnitudes and nonzero error correlations, these under and overestimations can result in negative estimates of error variance.

5. Truth and representativeness errors

In this section we discuss the relationship of the dataset of true values—truth—to differences in representativeness between datasets in the 3CH method, expanding upon the discussion of O'Carroll et al. (2008). The equations for error

FIG. 9. (left) The effect of unknown error correlations among two datasets on exact (solid) and estimated (dashed) normalized specific humidity error standard deviation (%). (right) Vertical profiles of normalized error covariance terms (%²) from this experiment. The correlation coefficients r_{YZ} of errors between Y and Z are (top) 0.1, (middle) 0.2, and (bottom) 0.4.

variances of three datasets X, Y, and Z given in Eqs. (3) are complete and exact no matter what truth is assumed, so long as it is the same for each dataset. Application of the 3CH method requires us to assume that the error covariance terms are zero or at least negligible compared to the mean square differences and bias terms. As O'Carroll et al. (2008) point out, the validity of this approximation depends on the definition of truth

through the error covariance terms. Changing the definition of truth does not change the error variance estimates, but it does change both the unknown error covariance terms and the exact error variances with respect to the defined truth.

One way in which truth may differ between the datasets is in the case of representativeness differences between them. Consider temperature for example. Some in situ observations

FIG. 10. (left) The effect of unknown error correlations among all datasets on exact (solid) and 3CH estimated (dashed) normalized specific humidity error standard deviation (%). (right) Vertical profiles of normalized error covariance terms (%²) from this experiment. Errors in Z are correlated (top) with errors in both X and Y at $r_{XZ} = r_{YZ} = 0.1$, and (bottom) with errors in X at $r_{XZ} = 0.1$ and in Y at $r_{YZ} = 0.2$.

such as radiosondes represent the temperature of a very small volume of the atmosphere—essentially a point—at a specific time. Other datasets such as infrared or microwave measurements from satellites represent a volume average or footprint over several kilometers in the horizontal and perhaps a kilometer in the vertical. The temperature obtained from NWP models or reanalyses provide the temperature averaged over the horizontal and vertical volume of a model cell. Since each system represents a different truth, a comparison of even error-free datasets that represent different scales will show differences that are sometimes called representativeness errors.

a. Vertical representativeness errors in the 3CH method

To see the effect of vertical representativeness errors in the 3CH method, consider three error-free datasets: one high-resolution dataset Z, and two datasets X and Y with the same lower resolution. The lower-resolution datasets are vertical averages of Z, and are “error-free” in the sense that they represent exact vertical averages of the high-resolution dataset. Note that this implies that $X \equiv Y$. Since the bias terms in Eqs. (3) are all zero here and we neglect the error covariance

terms, the 3CH estimated error variances for these datasets are zero for X and Y, and $E[(X - Z)^2]$ {or equivalently $E[(Y - Z)^2]$ } for Z.

Here, the differences in representation between the datasets will manifest in the 3CH method as a nonzero error variance estimate for Z. What we assume as truth T determines the values of the exact error variances and covariances, which we calculate by assuming either of the two low-resolution datasets are identically equal to truth or the high-resolution dataset is.

1) FIRST CASE: $X \equiv Y \equiv T$; $Z \neq T$

In the first case, we assume truth T is equal to the low-resolution datasets X and Y. Differences between Z and T are representativeness errors, denoted by ε_Z :

$$X = T, \quad (13a)$$

$$Y = T, \quad (13b)$$

$$Z = T + \varepsilon_Z. \quad (13c)$$

The error variances of X and Y are zero by definition, and it is straightforward to calculate the error variance of Z as

FIG. 11. Creating negative error variance estimates. (left) Exact (solid) and 3CH estimated (dashed) normalized temperature error variance (%²). (right) Vertical profiles of normalized error covariance terms (%²) from this experiment. Errors in X are uncorrelated with errors in Y and Z . Errors in Y and Z are correlated, and errors in Z are twice the magnitude of those in X and Y . Thus, $E[\varepsilon_Y \varepsilon_Z]$ is positive and large (see right panel), and the neglect of this term in Eqs. (3) results in an overestimate of the error variance estimate (see left panel) for X (dashed blue) and an underestimate for Y and Z (dashed red and dashed yellow, respectively). For Y , the neglect of $E[\varepsilon_Y \varepsilon_Z]$ results in a negative estimate of the error variance.

$$\begin{aligned} \text{Var}_{\text{err}}[Z] &= E[(X - Z)^2] \\ &= E[(Y - Z)^2]. \end{aligned} \quad (14)$$

With the assumed truth equal to X and Y , these are the exact error variances and equal to the 3CH estimates above.

2) SECOND CASE: $Z \equiv T$; $X \equiv Y \neq T$

For the second case, we assume truth is equal to the high-resolution dataset Z . Differences between Z and both X and Y are representativeness errors ε_X and ε_Y , respectively, with $\varepsilon_X \equiv \varepsilon_Y$:

$$X = T + \varepsilon_X, \quad (15a)$$

$$Y = T + \varepsilon_Y, \quad (15b)$$

$$Z = T, \quad (15c)$$

from which it is straightforward to calculate

$$\text{Var}_{\text{err}}[X] = E[(X - Z)^2], \quad (16a)$$

$$\text{Var}_{\text{err}}[Y] = E[(Y - Z)^2]. \quad (16b)$$

Thus, the estimated error variances differ from the true error variances by $E[(X - Z)^2]$ (or equivalently, $E[(Y - Z)^2]$) in the estimates for X and Y , and by $-E[(X - Z)^2]$ in the estimate for Z .

The above two cases also hold if X and Y are identical high-resolution datasets and Z is the low-resolution dataset. So long as two of the three datasets have the same resolution—either higher or lower than that of the third—the estimates produced by the 3CH method associate truth with the two datasets that have the same scale of representativeness. It all depends on what we define as truth, in agreement with O’Carroll et al. (2008).

We note that in both cases, because $X \equiv Y$, we could simply compute the error variance of the datasets directly, i.e., the 3CH method is not needed. The point here is to show how the

3CH method behaves in simple situations and how the assumption of truth affects the accuracy of the 3CH method estimates by changing the error covariances.

b. Vertical representativeness errors of radio occultation observations

We illustrate how representativeness errors manifest in the 3CH method as discussed in general in the previous section by considering a set of observed RO profiles on a high-resolution (100 m) grid and assuming they are truth. This set consists of all RO profiles between 1 and 7 January 2008, including those from the COSMIC, the Challenging Minisatellite Payload for Geophysical Research and Application (CHAMP), and the *Meteorological Operational Satellite-A* (MetOp-A). There are approximately 10 000 profiles in this initial set. We reduce this set by only taking those profiles that penetrate to 1 km, and by only analyzing data from 1 km and above. We are then left with 5994 profiles at all levels.

We classify this unaltered dataset as Z , and construct both datasets X and Y at one of two lower resolutions by averaging the RO profiles over vertical layers of 500 or 1300 m. All datasets may now be considered perfect on the scale that they represent, and differences are due only to representation. We evaluate the 3CH method at the midpoint of the layers so that no interpolation is necessary.

In our first experiment, X and Y are constructed as 1300 m layer averages of Z . This layer is considerably thicker than in typical model datasets, which are typically 200–300 m in the lower troposphere, but we chose this scale to illustrate the impact of vertical representativeness errors in an extreme case. The 3CH estimated error variance profile for Z (the unaltered dataset) is shown by the dashed yellow profile in the left panel of Fig. 12. The maximum errors due to representativeness differences occur in the lower troposphere below 6 km where the small-scale variability of RO is greatest, and near the tropopause (16–21 km) where the high-resolution dataset resolves the sharper tropopause relative to the lower-resolution datasets. For these different scales,

FIG. 12. The effect of vertical representativeness. The 3CH estimated normalized error variances of RO refractivity ($\% ^2$) where one dataset (Z ; dashed yellow) is taken at the original high resolution, and two others (X and Y with $X = Y$; dashed blue and dashed red, respectively) are generated by taking layer averages with depths of (left) 1300 and (right) 500 m. Note that the estimates for X and Y overlap perfectly and are zero. The estimated errors for Z may be considered representativeness errors for a high-vertical resolution dataset when the two comparison datasets have the same very coarse vertical resolution (see left panel) or vertical resolution more typical of current models (see right panel). Application of a second-order, 2500 m filter to Z in the very coarse example (see left panel) results in much smaller representativeness errors (dot-dashed yellow).

the representativeness errors of RO could be an appreciable fraction of the estimated error variances, which typically range between $1\% ^2$ and $10\% ^2$ in the tropics (Anthes and Rieckh 2018). The estimated error variances for the 1300 m averaged datasets (X and Y with $X \equiv Y$) are zero, as discussed in the previous section.

Because the vertical resolution of the above coarse datasets is much larger than most model datasets, we repeat this experiment, but with a coarse grid calculated as 500 m layer averages. As expected, the representativeness errors are reduced from a maximum of about $0.9\% ^2$ to about $0.1\% ^2$ (right panel of Fig. 12). The latter constitute a small fraction of the total error variance using RO and model data.

The assignment of representativeness errors in the 3CH method becomes more complicated if all three datasets have different vertical resolutions (Fig. 13). In this case, the 3CH method estimates nonzero error variances for all three datasets, and the estimates for one of the datasets become negative. As discussed in section 4e, true error variances must always remain positive, but negative error variance estimates can occur when one of the datasets (blue profile in this case) has a smaller actual error variance than another dataset to which its errors are strongly correlated. Figure 14 compares the true error variances and covariances when truth is assumed to be either the low-resolution (1300 m) dataset or the high-resolution (100 m) dataset. Irrespective of the assumed truth, the exact error variance for the 1100 m dataset (blue profiles in the left panels of Fig. 14) is less than the sum of the neglected error covariance terms (right panels of Fig. 14).

Vertical smoothing or filtering of the high-resolution dataset(s) so that structures with scales below those of the coarse resolution dataset(s) are decreased reduces representativeness errors (Kitchen 1989; Kuo et al. 2004; Lohmann 2007). We illustrate this by filtering the high-resolution dataset in our first experiment using a second-order filter (Savitzky and Golay 1964) with a vertical window of 2500 m—roughly double the low resolution here.

With filtering, the estimated error variance is reduced to under $0.1\% ^2$ everywhere (cf. the dashed and dot-dashed yellow profiles in the left panel of Fig. 12). Thus proper filtering can be an effective way of reducing representativeness errors to magnitudes that are negligible compared to the estimated error variances.

6. Summary and conclusions

This paper reviews the three-cornered hat (3CH) method for estimating the error variances of three datasets. The 3CH method

FIG. 13. The effect of vertical representativeness for three different vertical resolutions. 3CH error variance estimates of RO refractivity ($\% ^2$) for triplets including a high-resolution dataset (100 m, dashed yellow) and two lower-resolution datasets (1100 m, dashed blue; 1300 m, dashed red). The error variance of the intermediate-resolution dataset is negative.

FIG. 14. (left) True error variances and (right) error covariances for the experiment shown in Fig. 13 when truth is assumed to be either (top) the highest-resolution dataset or (bottom) the lowest-resolution dataset. In the left panels, the exact error variance of the dataset that corresponds to the selected truth is identically zero: 100 and 1300 m on top and bottom, respectively. In the top-right panel, the error covariances between the 100 m dataset and both the 1100 and 1300 m datasets (green and orange) are identically zero. In the bottom-right panel, the error covariances between the 1300 m dataset and both the 100 and 1100 m datasets (purple and orange) are identically zero.

is a special case of the more general N-cornered hat method, which uses N datasets. We show that there is a long scientific heritage of the method outside of the geophysical literature. A closely related method, called the triple collocation method, was developed independently and has been widely used to estimate the errors of three geophysical datasets. We have summarized these methods and their similarities and differences.

We generate synthetic datasets to analyze the factors that limit the accuracy of the 3CH method in real-world applications. These factors include sample size, outliers, different magnitudes of errors between the datasets, biases, and unknown error correlations. We find that biases and outliers are correctly accounted for by the method. Care should be taken to quality control data prior to applying the 3CH method, but the method will not find inaccurate error variance estimates for the input data due to biases or outliers.

For the assumed error models, we find that finite sampling leads to small nonzero error covariances that occur by chance. These finite sample size inaccuracies are exacerbated when the

magnitudes of error variance between the datasets is one or more orders of magnitude different. However, large sample sizes reduce these inaccuracies.

Error correlations between datasets represent a limitation of the method. Yet, we show that even for a relatively large error correlation of 0.2 between two datasets, the 3CH error variance estimates correctly represent the magnitude of the errors. Under the conditions of relatively small errors in one dataset and/or large error correlations between datasets, the 3CH error variance estimate may be negative. This is unphysical, and can thus be an effective identifier for triplets of datasets for which the 3CH method should not be used.

An additional condition under which the 3CH method may be inaccurate is when it is applied to datasets that have different representation. We show that even for three datasets that are error-free on three different vertical scales, the 3CH error variance estimates will not be zero on account of neglected error covariance terms. The effect of representativeness differences on error variance estimates may be reduced by applying appropriate scale filtering to the highest resolution dataset prior to use in the 3CH equations.

Under the conditions of three datasets with relatively large sampling (e.g., greater than 5000 points), similar scales of representation, similar magnitudes of error between them, and small error correlations among the data, we find that the 3CH method is an effective tool for accurately estimating the random error variances of all three datasets. These conditions are not overly restrictive and can allow for a wide range of applications of the method to understanding errors in geophysical data.

Acknowledgments. This work was supported by NSF–NASA Grant AGS-1522830 and NOAA Contract 16CN0070. We thank Eric DeWeaver (NSF) and Jack Kaye (NASA) for their long-term support of COSMIC. Data were retrieved from the COSMIC Data Analysis and Archive Center (<https://cdaac-www.cosmic.ucar.edu/>) and the Research Data Archive at the National Center for Atmospheric Research (<https://rda.ucar.edu/>).

APPENDIX A

Three- and N-Cornered Hat Derivations

Consider the set of datasets $\mathcal{X} = \{X1, X2, X3, \dots, XN\}$, where

$$\begin{aligned} X1 &= T + \beta_{X1} + \varepsilon_{X1}, \\ X2 &= T + \beta_{X2} + \varepsilon_{X2}, \\ X3 &= T + \beta_{X3} + \varepsilon_{X3}, \\ &\vdots \\ XN &= T + \beta_{XN} + \varepsilon_{XN}. \end{aligned} \quad (A1)$$

with T the set of unknown true values, or “truth;” β terms the mean bias terms relative to T ; and ε terms the the sets of

random variations, or errors, relative to T . Taking the first three elements of $\mathcal{X}$, the unique variance of differences are

$$\begin{aligned} \text{Var}[X1 - X2] &= \text{Var}[\varepsilon_{X1}] + \text{Var}[\varepsilon_{X2}] - 2E[\varepsilon_{X1}\varepsilon_{X2}], \\ \text{Var}[X1 - X3] &= \text{Var}[\varepsilon_{X1}] + \text{Var}[\varepsilon_{X3}] - 2E[\varepsilon_{X1}\varepsilon_{X3}], \\ \text{Var}[X2 - X3] &= \text{Var}[\varepsilon_{X2}] + \text{Var}[\varepsilon_{X3}] - 2E[\varepsilon_{X2}\varepsilon_{X3}], \end{aligned} \quad (A2)$$

where $E[\cdot]$ is the sample mean. With respect to the error variance terms $\text{Var}[\varepsilon]$, there are three equations and three unknowns that may be solved as, e.g.,

$$\begin{aligned} \text{Var}[\varepsilon_{X1}] &= \frac{1}{2}(\text{Var}[X1 - X2] + \text{Var}[X1 - X3] - \text{Var}[X2 - X3]) \\ &\quad + E[\varepsilon_{X1}\varepsilon_{X2}] + E[\varepsilon_{X1}\varepsilon_{X3}] - E[\varepsilon_{X2}\varepsilon_{X3}], \end{aligned} \quad (A3)$$

which are the 3CH error variance relations. In the text, we define error variance as, e.g., $\text{Var}_{\text{err}}[X1] \equiv \text{Var}[\varepsilon_{X1}]$. As well, note that we may split the variance of difference terms into their mean square difference and difference of bias forms as given in the text; e.g.,

$$\begin{aligned} \text{Var}[X1 - X2] &= E[X1 - X2]^2 - E[X1 - X2]^2 \\ &= E[(X1 - X2)^2] - b_{X1,X2}^2. \end{aligned} \quad (A4)$$

Taking the first four elements of $\mathcal{X}$, there are four equations and four unknown error variance terms. This holds for all N members of $\mathcal{X}$, and by induction, one may uniquely solve for any N error variance terms as

$$\begin{aligned} \text{Var}[\varepsilon_{Xi}] &= \frac{1}{N-1} \sum_{j=1}^N (1 - \delta_{ij}) \text{Var}[Xi - Xj] \\ &\quad - \frac{1}{(N-1)(N-2)} \sum_{j=1}^N \sum_{k=j}^N (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ik}) \text{Var}[Xj - Xk], \\ &\quad + \frac{2}{N-1} \sum_{j=1}^N (1 - \delta_{ij}) E[\varepsilon_{Xi}\varepsilon_{Xj}] \\ &\quad - \frac{2}{(N-1)(N-2)} \sum_{j=1}^N \sum_{k=j}^N (1 - \delta_{ij})(1 - \delta_{jk})(1 - \delta_{ik}) E[\varepsilon_{Xj}\varepsilon_{Xk}], \end{aligned} \quad (A5)$$

where δ is the Kronecker delta: $\delta_{ij} = 1$ for $i = j$ and 0 otherwise. These are the so-called N-cornered hat error variance relations, of which the 3CH ($N = 3$) is the simplest case.

Note that a consequence of this induction is that Eq. (A5) is equal to the mean of all unique, lower-order error variance relations. For example, for dataset $X1$, with $N = 4$ and no error covariance, Eq. (A5) expands to

$$\begin{aligned} \text{Var}[\varepsilon_{X1}] &= \frac{1}{3}(\text{Var}[X1 - X2] + \text{Var}[X1 - X3] + \text{Var}[X1 - X4]) \\ &\quad - \frac{1}{6}(\text{Var}[X2 - X3] + \text{Var}[X2 - X4] + \text{Var}[X3 - X4]). \end{aligned} \quad (A6)$$

The set of unique 3CH relations for this example are

$$\begin{aligned}\text{Var}[\varepsilon_{X1}] &= \frac{1}{2}(\text{Var}[X1 - X2] + \text{Var}[X1 - X3] - \text{Var}[X2 - X3]), \\ \text{Var}[\varepsilon_{X1}] &= \frac{1}{2}(\text{Var}[X1 - X2] + \text{Var}[X1 - X4] - \text{Var}[X2 - X4]), \\ \text{Var}[\varepsilon_{v1}] &= \frac{1}{2}(\text{Var}[X1 - X3] + \text{Var}[X1 - X4] - \text{Var}[X3 - X4]),\end{aligned}\quad (\text{A7})$$

the mean of which is equivalent to Eq. (A6). This holds for all N and all lower orders down to three, i.e., the 3CH. For any N datasets, there are $(N - 1)$ variance of difference relations that contain a given $\text{Var}[\varepsilon]$. The total number of unique 3CH relations for a given dataset are therefore given by “ $(N - 1)$ choose 2” or $(N - 1)(N - 2)/2$.

APPENDIX B

Three-Cornered Hat with Multiplicative Biases

As in appendix A, consider the set of datasets $\mathcal{X} = \{X1, X2, X3, \dots, XN\}$, where now

$$\begin{aligned}X1 &= \alpha_{X1}T + \beta_{X1} + \varepsilon_{X1}, \\ X2 &= \alpha_{X2}T + \beta_{X2} + \varepsilon_{X2}, \\ X3 &= \alpha_{X3}T + \beta_{X3} + \varepsilon_{X3}, \\ &\vdots \\ XN &= \alpha_{XN}T + \beta_{XN} + \varepsilon_{XN}.\end{aligned}\quad (\text{B1})$$

Here, α represent multiplicative bias terms for which $\alpha \neq 1$ implies that a given dataset is not calibrated. Following the derivation of appendix A, we may derive the following 3CH error variance equation for dataset X1:

$$\begin{aligned}\text{Var}[\varepsilon_{X1}] &= \frac{1}{2}(\text{Var}[X1 - X2] + \text{Var}[X1 - X3] - \text{Var}[X2 - X3]) \\ &\quad + E[\varepsilon_{X1}\varepsilon_{X2}] + E[\varepsilon_{X1}\varepsilon_{X3}] - E[\varepsilon_{X2}\varepsilon_{X3}] \\ &\quad - \alpha_{X1}\left(\alpha_{X1} - \alpha_{X2} - \alpha_{X3} + \frac{\alpha_{X2}\alpha_{X3}}{\alpha_{X1}}\right)E[T^2] \\ &\quad - (\alpha_{X1} - \alpha_{X2})E[T'(\varepsilon_{X1} - \varepsilon_{X2})] - (\alpha_{X1} - \alpha_{X3})E[T'(\varepsilon_{X1} - \varepsilon_{X3})] \\ &\quad + (\alpha_{X2} - \alpha_{X3})E[T'(\varepsilon_{X2} - \varepsilon_{X3})],\end{aligned}\quad (\text{B2})$$

where $T' = T - E[T]$ is the anomaly of truth T from its mean.

REFERENCES

- Anderson, W. B., B. F. Zaitchik, C. R. Hain, M. C. Anderson, M. T. Yilmaz, J. Mecikalski, and L. Schultz, 2012: Towards an integrated soil moisture drought monitor for East Africa. *Hydrol. Earth Syst. Sci.*, **16**, 2893–2913, <https://doi.org/10.5194/hess-16-2893-2012>.
- 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>.
- Barnes, J. A., 1966: Atomic timekeeping and the statistics of precision signal generators. *Proc. IEEE*, **54**, 207–220, <https://doi.org/10.1109/PROC.1966.4633>.
- Berrisford, P., and Coauthors, 2011: The ERA-Interim Archive: Version 2.0. ERA Rep. Series 1, 23 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2011/8174-era-interim-archive-version-20.pdf>.
- Blackmore, T. A., A. G. O’Carroll, R. W. Saunders, and H. H. Aumann, 2007: Numerical weather prediction: A comparison of sea surface temperature from the AATSR and AIRS instruments. *Met Office Tech. Rep.* 499, 21 pp.
- Bormann, N., M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally, 2016: Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quart. J. Roy. Meteor. Soc.*, **142**, 1767–1780, <https://doi.org/10.1002/qj.2774>.
- Braun, J., C. Rocken, and R. Ware, 2001: Validation of line-of-sight water vapor measurements with GPS. *Radio Sci.*, **36**, 459–472, <https://doi.org/10.1029/2000RS002353>.
- Caires, S., and A. Sterl, 2003: Validation of ocean wind and wave data using triple collocation. *J. Geophys. Res.*, **108**, 3098, <https://doi.org/10.1029/2002JC001491>.
- Chen, S.-Y., C.-Y. Huang, Y.-H. Kuo, and S. Sokolovskiy, 2011: Observational error estimation of FORMOSAT-3/COSMIC GPS radio occultation data. *Mon. Wea. Rev.*, **139**, 853–865, <https://doi.org/10.1175/2010MWR3260.1>.
- Crow, W. T., and M. J. van den Berg, 2010: An improved approach for estimating observation and model error parameters in soil moisture data assimilation. *Water Resour. Res.*, **46**, W12519, <https://doi.org/10.1029/2010WR009402>.
- Daley, R., 1993: Estimating observation-error statistics for atmospheric data assimilation. *Ann. Geophys.*, **11**, 634–647.
- Dee, D. P., and A. M. da Silva, 2003: The choice of variable for atmospheric moisture analysis. *Mon. Wea. Rev.*, **131**, 155–171, [https://doi.org/10.1175/1520-0493\(2003\)131<0155:TCOVFA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0155:TCOVFA>2.0.CO;2).
- , and Coauthors, 2011: 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>.
- 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>.
- , L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, <https://doi.org/10.1256/qj.05.108>.
- D’Odorico, P. A., A. Gonsamo, B. Pinty, N. Gobron, N. Coops, E. Mendez, and M. E. Schaepman, 2014: Intercomparison of fraction of absorbed photosynthetically active radiation products derived from satellite data over Europe. *Remote Sens. Environ.*, **142**, 141–154, <https://doi.org/10.1016/j.rse.2013.12.005>.
- Dorigo, W. A., K. Scipal, R. M. Parinussa, Y. Y. Liu, W. Wagner, R. A. M. de Jeu, and V. Naeimi, 2010: Error characterisation of global active and passive microwave soil moisture data sets. *Hydrol. Earth Syst. Sci.*, **14**, 2605–2616, <https://doi.org/10.5194/hess-14-2605-2010>.
- Draper C., R. Reichle, R. de Jeu, V. Naeimi, R. Parinussa, and W. Wagner, 2013: Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sens. Environ.*, **137**, 288–298, <https://doi.org/10.1016/j.rse.2013.06.013>.

- Ekstrom C. R., and P. A. Koppang, 2006: Error bars for three-cornered hats. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **53**, 876–879, <https://doi.org/10.1109/TUFFC.2006.1632679>.
- Fang, H., S. Wei, C. Jiang, and K. Scipal, 2012: Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method. *Remote Sens. Environ.*, **124**, 610–621, <https://doi.org/10.1016/j.rse.2012.06.013>.
- Galindo, F. J., J. J. Ruiz, E. Giachino, A. Premoli, and P. Tavella, 2001: Estimation of the covariance matrix of individual standards by means of comparison measurements. *Advanced Mathematical and Computational Tools in Metrology V*, World Scientific, 177–184, https://doi.org/10.1142/9789812811684_0020.
- Gentemann, C. L., 2014: Three way validation of MODIS and AMSR-E sea surface temperatures. *J. Geophys. Res. Oceans*, **119**, 2583–2598, <https://doi.org/10.1002/2013JC009716>.
- Gray, J. E., and D. W. Allan, 1974: A method for estimating the frequency stability of an individual oscillator. *Proc. 28th Frequency Control Symp.*, Atlantic City, NJ, IEEE, 243–246, <https://doi.org/10.1109/FREQ.1974.200027>.
- Greenhall, C. A., 1987: Likelihood and least squares approaches to the M-cornered hat. *Proc. 19th Annual Precise Time and Time Interval Applications and Planning Meeting*, Redondo Beach, CA, ION, 219–225.
- Griggs, E., E. R. Kursinski, and D. Akos, 2014: An investigation of GNSS atomic clock behaviour at short time intervals. *GPS Solutions*, **18**, 443–452, <https://doi.org/10.1007/s10291-013-0343-7>.
- , —, and —, 2015: Short-term GNSS satellite clock stability. *Radio Sci.*, **50**, 813–826, <https://doi.org/10.1002/2015RS005667>.
- Grubbs, F. E., 1948: On estimating precision of measuring instruments and product variability. *J. Amer. Stat. Assoc.*, **43**, 243–264, <https://doi.org/10.1080/01621459.1948.10483261>.
- Gruber, A., C.-H. Su, S. Zwieback, W. T. Crow, W. Dorigo, and W. Wagner, 2016: Recent advances in (soil moisture) triple collocation analysis. *Int. J. Appl. Earth Obs. Geoinf.*, **45**, 200–211, <https://doi.org/10.1016/j.jag.2015.09.002>.
- Hain, C. R., W. T. Crow, J. R. Mecikalski, M. C. Anderson, and T. Holmes, 2011: An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling. *J. Geophys. Res.*, **116**, D15107, <https://doi.org/10.1029/2011JD015633>.
- Janssen, P. A. E. M., S. Abdalla, H. Hersbach, and J.-R. Bidlot, 2007: Error estimation of buoy, satellite, and model wave height data. *J. Atmos. Oceanic Technol.*, **24**, 1665–1677, <https://doi.org/10.1175/JTECH2069.1>.
- Kitchen, M., 1989: Representativeness errors for radiosonde observations. *Quart. J. Roy. Meteor. Soc.*, **115**, 673–700, <https://doi.org/10.1002/qj.49711548713>.
- Kuo, Y.-H., T.-K. Wee, S. Sokolovskiy, C. Rocken, W. Schreiner, D. Hunt, and R. A. Anthes, 2004: Inversion and error estimation of GPS radio occultation data. *J. Meteor. Soc. Japan*, **82**, 507–531, <https://doi.org/10.2151/jmsj.2004.507>.
- Levine, J., 1999: Introduction to time and frequency metrology. *Rev. Sci. Instrum.*, **70**, 2567–2596, <https://doi.org/10.1063/1.1149844>.
- Loew, A., and F. Schlenz, 2011: A dynamic approach for evaluating coarse scale satellite soil moisture products. *Hydrol. Earth Syst. Sci.*, **15**, 75–90, <https://doi.org/10.5194/hess-15-75-2011>.
- Lohmann, M., 2007: Analysis of global positioning system (GPS) radio occultation measurement errors based on Satellite de Aplicaciones Científicas-C (SAC-C) GPS radio occultation data recorded in open-loop and phase-locked-loop mode. *J. Geophys. Res.*, **112**, D09115, <https://doi.org/10.1029/2006JD007764>.
- McColl, K. A., J. Vogelzang, A. G. Konings, D. Entekhabi, M. Piles, and A. Stoffelen, 2014: 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>.
- Miralles, D. G., W. T. Crow, and M. H. Cosh, 2010: Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *J. Hydrometeorol.*, **11**, 1423–1429, <https://doi.org/10.1175/2010JHM1285.1>.
- O’Carroll, A. G., J. R. Eyre, and R. W. Saunders, 2008: Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *J. Atmos. Oceanic Technol.*, **25**, 1197–1207, <https://doi.org/10.1175/2007JTECH0542.1>.
- Parinussa, R. M., T. R. H. Holmes, M. T. Yilmaz, and W. T. Crow, 2011: The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. *Hydrol. Earth Syst. Sci.*, **15**, 3135–3151, <https://doi.org/10.5194/hess-15-3135-2011>.
- Portabella, M., and A. Stoffelen, 2009: On scatterometer ocean stress. *J. Atmos. Oceanic Technol.*, **26**, 368–382, <https://doi.org/10.1175/2008JTECH0578.1>.
- Premoli, A., and P. Tavella, 1993: A revisited three-cornered hat method for estimating frequency standard stability. *IEEE Trans. Instrum. Meas.*, **42**, 7–13, <https://doi.org/10.1109/19.206671>.
- 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>.
- Riley, W. J., 2008: Handbook of frequency stability analysis. NIST Special Publ. 1065, 136 pp., <https://www.nist.gov/publications/handbook-frequency-stability-analysis>.
- Roebeling, R. A., E. L. A. Wolters, J. F. Meirink, and H. Leijnse, 2012: Triple collocation of summer precipitation retrievals from SEVIRI over Europe with gridded rain gauge and weather radar data. *J. Hydrometeorol.*, **13**, 1552–1566, <https://doi.org/10.1175/JHM-D-11-089.1>.
- Savitzky, A., and M. J. E. Golay, 1964: Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, **36**, 1627–1639, <https://doi.org/10.1021/ac60214a047>.
- Schröder, M., and Coauthors, 2018: The GEWEX Water Vapor Assessment archive of water vapour products from satellite observations and reanalyses. *Earth Syst. Sci. Data*, **10**, 1093–1117, <https://doi.org/10.5194/essd-10-1093-2018>.
- Scipal, K., T. Holmes, R. de Jeu, V. Naeimi, and W. Wagner, 2008: A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophys. Res. Lett.*, **35**, L24403, <https://doi.org/10.1029/2008GL035599>.
- Simmons, A. J., and Coauthors, 2016: Observation and integrated Earth-system science: A roadmap for 2016–2025. *Adv. Space Res.*, **57**, 2037–2103, <https://doi.org/10.1016/j.asr.2016.03.008>.
- Smith, E., and S. Weintraub, 1953: The constants in the equation for atmospheric refractive index at radio frequencies. *Proc. IRE*, **41**, 1035–1037, <https://doi.org/10.1109/JRPROC.1953.274297>.
- Stoffelen, A., 1998: 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>.
- Su, C.-H., D. Ryu, W. T. Crow, and A. W. Western, 2014: Beyond triple collocation: Applications to soil moisture monitoring. *J. Geophys. Res. Atmos.*, **119**, 6419–6439, <https://doi.org/10.1002/2013JD021043>.

- Tavella, P., and A. Premoli, 1994: Estimating the instabilities of N clocks by measuring differences of their readings. *Metrologia*, **30**, 479–486, <https://doi.org/10.1088/0026-1394/30/5/003>.
- Vogelzang, J., and A. Stoffelen, 2012: Triple collocation. EUMETSAT Tech. Rep. NWPSAF-KN-TR-021, 22 pp., https://www.nwpsaf.eu/site/download/documentation/scatterometer/TripleCollocation_NWPSAF_TR_KN_021_v1_0.pdf.
- , ———, A. Verhoef, and J. Figa-Saldaña, 2011: On the quality of high-resolution scatterometer winds. *J. Geophys. Res.*, **116**, C10033, <https://doi.org/10.1029/2010JC006640>.
- Weston, P. P., W. Bell, and J. R. Eyre, 2014: Accounting for correlated error in the assimilation of high-resolution sounder data. *Quart. J. Roy. Meteor. Soc.*, **140**, 2420–2429, <https://doi.org/10.1002/qj.2306>.
- Yilmaz, M. T., and W. T. Crow, 2013: The optimality of potential rescaling approaches in land data assimilation. *J. Hydrometeor.*, **14**, 650–660, <https://doi.org/10.1175/JHM-D-12-052.1>.
- Zwieback, S., K. Scipal, W. Dorigo, and W. Wagner, 2012: Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Processes Geophys.*, **19**, 69–80, <https://doi.org/10.5194/npg-19-69-2012>.