The N-Cornered Hat method for Estimating Error Variances between Multiple Data Sets:

Theoretical considerations and comparisons with the two-cornered hat method

Jeremiah P Sjoberg, Richard A Anthes, and Therese Rieckh
UCAR/COSMIC

Abstract

Variations of the three-cornered hat (3CH) method for estimating random error variances associated with three or more data sets have been reported in the literature for studying geophysical data sets, including sea-surface winds, sea-surface temperatures, precipitation, leaf area index and soil moisture. Anthes and Rieckh (2018) and Rieckh and Anthes (2018) used the 3CH method to estimate the error variances of multiple atmospheric sounding data sets. However, the methods used often contain subtle variations and make different assumptions. Here we derive the full 3CH equations that relate the error variance of three or more observations to the variance of differences between the data sets and the error covariances among the different data sets.

Three-cornered hat relations

Consider the data sets $X_1, X_2$, and $X_3$ with individual elements $i = 1, 2, \ldots$. We assume that these may be cast as

\[ X_{1i} = T_i + \epsilon_{11} + \epsilon_{12}, \]
\[ X_{2i} = T_i + \epsilon_{21} + \epsilon_{22}, \]
\[ X_{3i} = T_i + \epsilon_{31} + \epsilon_{32}, \]

where $T_i$ is a set of reference values, the $\epsilon$ terms are mean differences of individual data sets from the reference data set, and the $\epsilon$ terms are sets of zero mean, not necessarily Gaussian random variations.

To remove the mean difference terms, we subtract the mean $E[\cdot]$ of each data set:

\[ X'_{1i} = \epsilon_{11} + \epsilon_{12}, \]
\[ X'_{2i} = \epsilon_{21} + \epsilon_{22}, \]
\[ X'_{3i} = \epsilon_{31} + \epsilon_{32}, \]

where primes denote difference from the mean.

The unique set of the variance of differences between data sets can be written

\[ \text{Var}(X_{1i} - X_{2i}) = \text{Var}(\epsilon_{11}) + \text{Var}(\epsilon_{21}) + 2 \text{Cov}(\epsilon_{11}, \epsilon_{21}) \]
\[ \text{Var}(X_{1i} - X_{3i}) = \text{Var}(\epsilon_{11}) + \text{Var}(\epsilon_{31}) + 2 \text{Cov}(\epsilon_{11}, \epsilon_{31}) \]
\[ \text{Var}(X_{2i} - X_{3i}) = \text{Var}(\epsilon_{21}) + \text{Var}(\epsilon_{31}) + 2 \text{Cov}(\epsilon_{21}, \epsilon_{31}) \]
\[ \text{Cov}(\epsilon_{11}, \epsilon_{21}) + \text{Cov}(\epsilon_{11}, \epsilon_{31}) + \text{Cov}(\epsilon_{21}, \epsilon_{31}) \]

where $\text{Cov}(\cdot)$ is the covariance between two quantities.

The relations for error variance can be derived by linearly combining Eqs. (3a-c):

\[ \text{Var}(\epsilon_{11}) = \frac{1}{2}(\text{Var}(X_{1i} - X_{2i}) + \text{Var}(X_{1i} - X_{3i}) - \text{Var}(X_{2i} - X_{3i})) \]
\[ \text{Var}(\epsilon_{21}) = \frac{1}{2}(\text{Var}(X_{1i} - X_{2i}) + \text{Var}(X_{2i} - X_{3i}) - \text{Var}(X_{1i} - X_{3i})) \]
\[ \text{Var}(\epsilon_{31}) = \frac{1}{2}(\text{Var}(X_{2i} - X_{3i}) + \text{Var}(X_{1i} - X_{3i}) - \text{Var}(X_{1i} - X_{3i})) \]

Four-cornered hat relations

Following the derivation of the 3CH relations, it may be shown that the four-cornered hat error variance relationship for data set $X_4$ is

\[ \text{Var}(\epsilon_{41}) = \frac{1}{2}(\text{Var}(X_{1i} - X_{2i}) + \text{Var}(X_{1i} - X_{3i}) + \text{Var}(X_{1i} - X_{4i})) \]
\[ \text{Var}(\epsilon_{42}) = \frac{1}{2}(\text{Var}(X_{2i} - X_{3i}) + \text{Var}(X_{2i} - X_{4i}) + \text{Var}(X_{1i} - X_{4i})) \]
\[ \text{Var}(\epsilon_{43}) = \frac{1}{2}(\text{Var}(X_{3i} - X_{4i}) + \text{Var}(X_{3i} - X_{4i}) + \text{Var}(X_{1i} - X_{4i})) \]

Other variance relationships

The set of variances for $X_{ij}$ can be written

\[ \text{Var}(X_{ij}) = \text{Var}(\epsilon_{ij}) + \text{Var}(T_i + \epsilon_{ij} + \epsilon_{ij} + \epsilon_{ij}) \]
\[ = \text{Var}(\epsilon_{ij}) + \text{Var}(T_i) + 2E[\epsilon_{ij} + \epsilon_{ij} + \epsilon_{ij} + \epsilon_{ij}] \]
\[ = E[\text{Var}(\epsilon_{ij})] + 2E[\epsilon_{ij} + \epsilon_{ij} + \epsilon_{ij} + \epsilon_{ij}] \]
\[ = \text{Var}(T_i) + 2\text{Var}(\epsilon_{ij}) + 4\text{Cov}(\epsilon_{ij}, \epsilon_{ij}) \]

For data sets $X_i$ and $X_j$, the variance of the sum is

\[ \text{Var}(X_i + X_j) = \text{Var}(T_i + T_j + \epsilon_{ij} + \epsilon_{ij}) \]

Summary

For data sets that can be cast following Eqs. (1a-c), we may derive the 3CH relations for error variance by assuming that there is no error covariance between the three data sets, we may apply this method to observations.

Here we show that:

- The 3CH method can be generalized for error variance estimation using $N$ different data sets.
- The single estimate of error variance using the $N$ data sets is equal to the mean of the $N(N-1)/2$ estimates of error variances for each individual data set using the 3CH method. These estimates can be used to evaluate our assumption of no error covariance between any given triplet of data sets.
- The 3CH method contains as a subset three estimates of error variance using only two sets of data — sometimes called the “triple co-location method.” Application of this method requires making assumptions about the reference data set that will lead to decreased accuracy.

References

- Anthes and Rieckh (2018), DOI: 10.5194/amt-11-4309-2018
- Rieckh and Anthes (2018), DOI: 10.5194/amt-11-4309-2018

Contact*

sjoberg@ucar.edu

Example: 3CH with six data sets

The two-cornered hat method

The two-cornered hat or “triple co-location method” — error variance relation for $X_{ij}$ is found by taking Eqs. (9)-(12a)-4(7a):