RESEARCH ARTICLE

The Relationship Between Two Methods for Estimating Uncertainties in Data Assimilation

Ricardo Todling[1] | Noureddine Semane[2] | Richard Anthes[3] | Sean Healy[2]

[1]NASA Global Modeling and Assimilation Office, Greenbelt MD, USA
[2]European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
[3]COSMIC Program Office, University Corporation for Atmospheric Research, Boulder, Colorado, USA

Correspondence
Ricardo Todling, NASA Global Modeling and Assimilation Office, Code 610.1, Greenbelt MD, 20771, USA
Email: ricardo.todling@nasa.gov

Funding information
Todling: NASA GMAO Core Funding; Semane: Radio Occultation Meteorology Satellite Application Facility (ROM SAF; EUMETSAT); Anthes: NSF Grant AGS-2054356 and NOAA Contract 16CN0070

This note examines the relationship between two apparently unrelated methods for estimating error statistics or uncertainties of relevance to data assimilation. The first method is due to (Desroziers et al., 2005, Q. J. R. Meteorol. Soc., 131, 3385–3396; referred to as DBCP hereafter) and relies on residual statistics readily available from data assimilation applications. The second method, the three-cornered hat (3CH) developed by Gray and Allan (1974, IEEE 28th Annual Symp. Freq. Control, 243–246), only recently applied to atmospheric sciences, uses three data sets and can derive estimates of relevant error uncertainties as well. The usefulness of both methods lies in them not requiring knowledge of the true value of the quantities at play. DBCP derives its results by relying explicitly on the constraints associated with the data assimilation minimization problem; 3CH is general and its estimates hold as long as errors in the three data sets of choice are uncorrelated. Establishing the relationship between the methods requires applying the 3CH approach to the same observation, background, and analysis data sets used by DBCP. In this case, the same assumptions of DBCP on residual errors allow for cancellation of error cross–covariance terms in 3CH such that two of its corners derive identical estimates for observation and background error covariances as those of DBCP. The error cross–covariance terms associated with the third corner are shown to add up to twice the analysis error covariance so that the 3CH estimate for the third corner recovers the negative of the analysis error covariance. Illustrations of these findings are provided by deriving uncertainties for radio occultation bending angles.

KEYWORDS
Residual statistics, three–cornered hat, Kalman filter.

1 | BRIEF BACKGROUND

Two methods for estimating error (co)variance that seem unrelated, and based on very different assumptions, are shown here to be very closely related and reproduce each others results in a special case. The methods in question are that introduced by Desroziers et al. (2005, DBCP hereafter), which relies on two sequences of observation residuals produced in typical data assimilation methods, and the three-cornered hat (3CH) method introduced by Gray and Allan (1974), which is developed in a general statistical context and relies on availability of three data sets providing information about the observable of interest. Both methods avoid the need for knowing the true value of the quantity of interest. The realization that a precise statement about how the two methods compare came about during the review process of the work of Semane et al. (2022) showing a *numerical* comparison of estimates derived by these two methods for radio occultation (RO) bending angle observations.

DBCP can be employed to derive estimates of observation, background and analysis error standard deviation (or variance) associated with given observables used in data assimilation (DA) applications. The method is frequently applied at NWP centers to check consistency and tune second order statistics required in corresponding DA systems. The particular case of estimating *observation* errors relies on constructing the following error covariance

$$\hat{\mathbf{R}}_{DBCP} = E\left[(\mathbf{o}-\mathbf{a})(\mathbf{o}-\mathbf{b})^T\right], \tag{1}$$

from *sample* data. The variable $\mathbf{o}$ represents a p-vector of observations, $\mathbf{b}$ and $\mathbf{a}$ represent background and analysis fields projected onto the p–dimensional space of observations by a suitable observation operator – the notation here intentionally hides such operator. The symbol $E[\bullet]$ represents the expectation operator; the difference vectors in the parenthesis are the so–called residual vectors. The variance (and corresponding standard deviation) can be extracted from the diagonal of this expression, that is, $\mathbf{\Sigma}_o^2 = diag(\hat{\mathbf{R}}_{DBCP})$. Given that DBCP provide expressions for full covariance matrices, the method has also been used extensively to extract relevant observation error correlations (e.g., Stewart et al. 2014; Weston et al. 2014; Bormann et al. 2016; Waller et al. (2016, 2019)).

The 3CH method chooses three data sets, $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, providing estimates of the quantity of interest and generates an estimate of the error covariance in data set $\{\mathcal{X}\}$ as in

$$
\begin{aligned}
\hat{\mathbf{X}} &= \frac{1}{2}\left\{E\left[(\mathbf{x}-\mathbf{y})(\mathbf{x}-\mathbf{y})^T\right]+E\left[(\mathbf{x}-\mathbf{z})(\mathbf{x}-\mathbf{z})^T\right]\right. \\
&\quad -E\left[(\mathbf{y}-\mathbf{z})(\mathbf{y}-\mathbf{z})^T\right] \\
&\quad \left. -\left(\mathbf{B}^{xy}+\mathbf{B}^{xz}-\mathbf{B}^{zy}\right)\right\} \\
&+ \left[E\left(\boldsymbol{\epsilon}^{x}\odot\boldsymbol{\epsilon}^{y}\right)+E\left(\boldsymbol{\epsilon}^{x}\odot\boldsymbol{\epsilon}^{z}\right)-E\left(\boldsymbol{\epsilon}^{z}\odot\boldsymbol{\epsilon}^{y}\right)\right] \\
&= \frac{1}{2}\left\{cov(\mathbf{x}-\mathbf{y})+cov(\mathbf{x}-\mathbf{z})-cov(\mathbf{y}-\mathbf{z})\right\} \\
&\quad +\left[E(\boldsymbol{\epsilon}^{x}\odot\boldsymbol{\epsilon}^{y})+E(\boldsymbol{\epsilon}^{x}\odot\boldsymbol{\epsilon}^{z})-E(\boldsymbol{\epsilon}^{y}\odot\boldsymbol{\epsilon}^{z})\right], \qquad (2)
\end{aligned}
$$

where $\mathbf{B}^{xy} \equiv E(\mathbf{x}-\mathbf{y})E(\mathbf{x}-\mathbf{y})^T$, $\mathbf{B}^{xz} \equiv E(\mathbf{x}-\mathbf{z})E(\mathbf{x}-\mathbf{z})^T$, and $\mathbf{B}^{yz} \equiv E(\mathbf{y}-\mathbf{z})E(\mathbf{y}-\mathbf{z})^T$, with errors defined as, $\boldsymbol{\epsilon}^{x} = \mathbf{x} - E[\mathbf{x}] - \mathbf{t}$, and analogously for $\boldsymbol{\epsilon}^{y}$ and $\boldsymbol{\epsilon}^{z}$; the last line in the first equality uses tensor notation to express $\mathbf{u}\odot\mathbf{v} = 1/2(\mathbf{u}\mathbf{v}^T+\mathbf{v}\mathbf{u}^T)$, for two arbitrary p-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$. The second equality uses the more compact definition of a covariance matrix: $cov(\mathbf{u},\mathbf{v}) = E[(\mathbf{u}-E(\mathbf{u}))(\mathbf{v}-E(\mathbf{v}))]^T = E(\mathbf{u}\mathbf{v}^T) - E(\mathbf{u})E(\mathbf{v}^T)$, with $cov(\mathbf{u}) = cov(\mathbf{u},\mathbf{u})$.

Practical applications of 3CH seek three data sets with independent errors, as to guarantee each of the cross–terms in (2) to be zero so they can be safely neglected. This can be accomplished by using independent observations, short–term model forecasts, or model data sets that do not assimilate the observations in the sample studied. As shown later in this note, it turns out that all that needs to happen is for there to be cancellation of the cross–covariance terms. The remaining terms in (2) are the only ones that can be calculated from sample data. The order of the data sets is arbitrary, which means that equivalent expressions for the error (co)variances $\mathbf{Y}$ and $\mathbf{Z}$ associated with data sets $\{\mathcal{Y},\mathcal{Z}\}$ can be obtained from (2) by rotating the variables $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$. Anthes and Rieckh (2018) introduced the first application of the 3CH method to atmospheric data sets. When using 3CH to derive estimates of observation error statistics the remaining two corners are arbitrary and can be chosen at will. It is plausible to wonder how DBCP observation error uncertainty estimates compare with those of 3CH. This is the motivation for the numerical work of Semane et al. (2022) looking at RO bending angle observations.

Given that analysis errors are dependent on observation and background errors, it seems puzzling to expect 3CH to derive anything reasonable for observation uncertainties when its remaining two corners are formed by the background and analysis. From the start, 3CH would seem to violate its requirement to work with data sets with uncorrelated errors. Therefore, the work of Semane et al. (2022) showing reasonable agreement between estimates from two of the corners of 3CH with DBCP estimates of observation and background errors needs further understanding. Beyond that, one might wonder what the remaining corner of 3CH obtains and how it relate with DBCP's estimate of analysis error. The present work finds that, under similar assumptions used to derive DBCP, two of the 3CH corners *exactly* recover DBCP error uncertainty estimates for the observations and backgrounds; while, surprisingly, its third corner recovers the negative of the error covariance associated with the analysis.

This work does not aim to provide a comprehensive review of the literature on either of the methods investigated here. Sjoberg et al. (2021) give a review of 3CH, its history, intricacies and limitations, including its relationship with the triple-collocation method of Stoffelen (1998). Tandeo et al. (2020) give a review of what the authors refer to as "innovation–based methods", but should more generally be referred to as "residual–based methods", of which DBCP is one example. This work confines itself with simply establishing the relationship between these methods. In what follows, Section 2 establishes the full relationship between a particular choice of corners for 3CH with the *optimal* estimates from DBCP; an Appendix establishes the relationship in the general suboptimal case. Section 3 provides standard deviation estimates of RO bending angle, revising similar results in Semane et al. (2022) in light of the relationship established in this work; an additional illustration is provided by comparing observation error correlation estimates derived by the two methods for RO bending angle.

2 | RELATIONSHIP BETWEEN DBCP AND 3CH

2.1 | Equivalence of 3CH and DBCP when estimating R

For simplicity, assume for now that no biases are at play. Subtracting the truth from the observation, analysis and background in a symmetrized form of (1) the DBCP estimate for the observation error covariance can be written as

$$\hat{\mathbf{R}} = E\left[(\boldsymbol{\epsilon}^{o} - \boldsymbol{\epsilon}^{a}) \odot (\boldsymbol{\epsilon}^{o} - \boldsymbol{\epsilon}^{b})\right], \tag{3}$$

and cross-multiplying the terms on the rhs of the expression above obtains

$$\mathbf{R} = \hat{\mathbf{R}} + E\left(\boldsymbol{\epsilon}^{o} \odot \boldsymbol{\epsilon}^{b}\right) + E\left(\boldsymbol{\epsilon}^{a} \odot \boldsymbol{\epsilon}^{o}\right) - E\left(\boldsymbol{\epsilon}^{a} \odot \boldsymbol{\epsilon}^{b}\right), \tag{4}$$

where $\mathbf{R} = E(\boldsymbol{\epsilon}^{o}\boldsymbol{\epsilon}^{o\,T})$ is the sought out observation error covariance matrix.

Now $\hat{\mathbf{R}}$ can also be expressed in at least two alternative ways. First, by adding and subtracting $\mathbf{a}$ to the second term in parenthesis on the rhs of (1) it follows that

$$\begin{aligned} \hat{\mathbf{R}} &= E\left\{[\mathbf{o} - \mathbf{a}] \odot [\mathbf{o} - \mathbf{a} - (\mathbf{b} - \mathbf{a})]\right\} \\ &= E\left[(\mathbf{o} - \mathbf{a}) \odot (\mathbf{o} - \mathbf{a})\right] - E\left[(\mathbf{o} - \mathbf{a}) \odot (\mathbf{b} - \mathbf{a})\right], \end{aligned} \tag{5}$$

and alternatively, by adding and subtracting $\mathbf{b}$ to the first term in parenthesis on the rhs of (1) it follows that

$$\begin{aligned} \hat{\mathbf{R}} &= E\left\{[(\mathbf{o} - \mathbf{b}) - (\mathbf{a} - \mathbf{b})] \odot (\mathbf{o} - \mathbf{b})\right\} \\ &= E\left[(\mathbf{o} - \mathbf{b}) \odot (\mathbf{o} - \mathbf{b})\right] - E\left[(\mathbf{a} - \mathbf{b}) \odot (\mathbf{o} - \mathbf{b})\right] \\ &. \end{aligned} \tag{6}$$

Adding (5) and (6), and reordering the terms a little,

$$\begin{aligned} \hat{\mathbf{R}} &= \frac{1}{2}\left\{E\left[(\mathbf{o} - \mathbf{a}) \odot (\mathbf{o} - \mathbf{a})\right] - E\left[(\mathbf{o} - \mathbf{a}) \odot (\mathbf{b} - \mathbf{a})\right] + \right. \\ &\qquad \left. E\left[(\mathbf{o} - \mathbf{b}) \odot (\mathbf{o} - \mathbf{b})\right] + E\left[(\mathbf{o} - \mathbf{b}) \odot (\mathbf{b} - \mathbf{a})\right]\right\} \\ &= \frac{1}{2}\left\{E\left[(\mathbf{o} - \mathbf{a}) \odot (\mathbf{o} - \mathbf{a})\right] - E\left[(\mathbf{a} - \mathbf{b}) \odot (\mathbf{a} - \mathbf{b})\right] + E\left[(\mathbf{o} - \mathbf{b}) \odot (\mathbf{o} - \mathbf{b})\right]\right\}. \end{aligned}$$

Substituting this result in (4) leads to

$$\begin{aligned} \mathbf{R} &= \frac{1}{2}\left\{cov(\mathbf{o} - \mathbf{b}) + cov(\mathbf{o} - \mathbf{a}) - cov(\mathbf{a} - \mathbf{b})\right\} \\ &+ \left[E\left(\boldsymbol{\epsilon}^{o} \odot \boldsymbol{\epsilon}^{b}\right) + E\left(\boldsymbol{\epsilon}^{a} \odot \boldsymbol{\epsilon}^{o}\right) - E\left(\boldsymbol{\epsilon}^{a} \odot \boldsymbol{\epsilon}^{b}\right)\right]. \end{aligned} \tag{7}$$

Let us now identify the three data sets, $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, associated with 3CH to be the observation, background and analysis, $\{\mathcal{O}, \mathcal{B}, \mathcal{A}\}$, i.e., take $\mathbf{o} = \mathbf{x}$, $\mathbf{b} = \mathbf{y}$, and $\mathbf{a} = \mathbf{z}$. With these, (7) derives (2), i.e., the DBCP result for $\mathbf{R}$ is identical to that of 3CH. However, once 3CH neglects the cross terms in (7), it would seem to no longer agree with DBCP. We will arrive at a full understanding of why 3CH still recovers DBCP even when these terms are neglected. Before that, we provide a brief recap of DBCP.

2.2 | Brief recap of DBCP: the optimal case

The result above derives the 3CH estimate for the observation error uncertainty from DBCP's. Alternatively, one can start from the 3CH general result for its three corners, apply the assumptions of DBCP, and see what derives. For that, it helps recapitulate the assumptions of DBCP. Although the point of the DBCP diagnostic is to inform on the statistics of errors in the general suboptimal case, in this section, for simplicity, we take DBCP's results in its optimal form. The general suboptimal case is treated in Appendix A.

We start by saying a word about the notation adopted in this article. Dimensions that typically refer to state–space are thought of as projections to observation space using a suitable observation operator, $\mathbf{H}$. With that, the vectors of background and analysis $\mathbf{b}$ and $\mathbf{a}$, appearing in expressions such as (1), and their corresponding errors, $\boldsymbol{\epsilon}^{\boldsymbol{b}}$ and $\boldsymbol{\epsilon}^{\boldsymbol{a}}$, are collapsed versions of what would normally be written as $\mathbf{Hx}^b$, $\mathbf{Hx}^a$, $\mathbf{He}^b$ and $\mathbf{He}^a$, with $\mathbf{x}^b$, $\mathbf{x}^a$, $\mathbf{e}^b$ and $\mathbf{e}^a$ being the full state-space background and analysis, and their respective errors. Similarly, $\mathbf{B}$ is the notation for what normally would be written as $\mathbf{HBH}^T$, and analogously for $\mathbf{A}$. The case of a matrix like the Kalman gain, whose only first dimension refers to the state–space, the notation implies that the $\mathbf{H}$ operator should appear on the left side of the original matrix, that is, $\mathbf{K}$ in this case stands for what would typically appear as $\mathbf{HK}$. Implicit in the notation is the assumption of linearity of the observation operator.

The assumptions required for DBCP are as follows:

Assumption-1: That observation and background errors be uncorrelated: $E(\boldsymbol{\epsilon}^{\boldsymbol{o}}\boldsymbol{\epsilon}^{\boldsymbol{b}^T}) = 0$.
Assumption-2: That analysis errors be linearly related with observation and background errors:

$$\boldsymbol{\epsilon}^{\boldsymbol{a}} = \boldsymbol{\epsilon}^{\boldsymbol{b}} + \mathbf{K}(\boldsymbol{\epsilon}^{\boldsymbol{o}} - \boldsymbol{\epsilon}^{\boldsymbol{b}})\,. \tag{8}$$

This last assumption is associated with the linearity of the observation operator mentioned above. To simplify the derivation of the relationship between 3CH and DBCP that follows, we take the trivial case when DBCP has the weighting matrix $\mathbf{K}$ set to be the Kalman gain, of *optimal* filtering theory,

$$\mathbf{K} = \mathbf{B}(\mathbf{B} + \mathbf{R})^{-1}\,, \tag{9}$$

where $\mathbf{B} = E(\boldsymbol{\epsilon}^{\boldsymbol{b}}\boldsymbol{\epsilon}^{\boldsymbol{b}^T})$ is the observation space projection of the background error covariance in the compact notation. We emphasize this choice is made here for convenience; DBCP is all about the suboptimality of the weighing matrix. It is also useful to point out that (8) and (9) imply that the analysis error is orthogonal to the innovation vector, $\mathbf{o} - \mathbf{b}$, that is, $E[\boldsymbol{\epsilon}^{\boldsymbol{a}}(\boldsymbol{\epsilon}^{\boldsymbol{o}} - \boldsymbol{\epsilon}^{\boldsymbol{b}})^T] = 0$; this result will be explicitly used later — it is the basis of linear estimation (Kalman 1960; see also Lewis et al. 2006, Chap. 6).

Remark 1: DBCP and bias. The theoretical derivation of DBCP assumes the observation residuals of the underlying DA system to be unbiased. In practice observation residuals are never fully unbiased, and construction of the cross–

covariance (1) is done by subtracting the residual biases. In this sense it is proper to write DBCP as

$$\hat{\mathbf{R}}_u \equiv cov(\mathbf{o}-\mathbf{a}, \mathbf{o}-\mathbf{b})\,. \tag{10}$$

Remark 2: DBCP as a covariance estimator. In actuality (10) represents a cross–covariance while it attempts to estimate a covariance. The two requirements for a matrix to be a covariance are symmetry and positive semi–definiteness. The first can be satisfied by introducing a symmetric version of DBCP,

$$\hat{\mathbf{R}} \equiv \frac{1}{2}(\hat{\mathbf{R}}_u + \hat{\mathbf{R}}_u^T)\,. \tag{11}$$

This is not unnatural, and practical use of DBCP that attempts to estimate *correlations* typically employs symmetrization (e.g.,Gauthier et al. 2018, Waller et al. 2019, Aabaribaoune et al. 2021, Cheng and Qiu 2021). The second requirement of positive semi–definiteness must be observed carefully when constructing covariances from finite samples. When it comes to using sample covariances in the algorithms of DA there is need for carefully ensuring positive–definiteness and avoiding poorly–conditioned matrices. A number of works have considered these matters in detail (viz., Weston et al. 2014; Geer 2019; Tabeart et al. 2020a,b). Symmetry is not an issue in 3CH. When it comes to positive semi–definiteness and conditioning, sample size and noise in the data affect 3CH just as much as DBCP.

Using Assumptions–1 and –2, DBCP derives the following expressions, under optimality of the gain matrix:

$$\hat{\mathbf{R}} \equiv \frac{1}{2}\left[cov(\mathbf{o}-\mathbf{a},\mathbf{o}-\mathbf{b}) + cov(\mathbf{o}-\mathbf{b},\mathbf{o}-\mathbf{a})\right] \overset{opt}{=} \mathbf{R}\,, \tag{12a}$$

$$\hat{\mathbf{B}} \equiv \frac{1}{2}\left[cov(\mathbf{a}-\mathbf{b},\mathbf{o}-\mathbf{b}) + cov(\mathbf{o}-\mathbf{b},\mathbf{a}-\mathbf{b})\right] \overset{opt}{=} \mathbf{B}\,, \tag{12b}$$

$$\hat{\mathbf{A}} \equiv \frac{1}{2}\left[cov(\mathbf{a}-\mathbf{b},\mathbf{o}-\mathbf{a}) + cov(\mathbf{o}-\mathbf{a},\mathbf{a}-\mathbf{b})\right] \overset{opt}{=} \mathbf{A}\,, \tag{12c}$$

where $\mathbf{A} \equiv (\mathbf{I}-\mathbf{K})\mathbf{B}$ is the analysis error covariance.

2.3 | Full relationship between 3CH and DBCP

With the index rotation mentioned in the introduction, the 3CH uncertainties associated with all three data sets $\{\mathcal{X},\mathcal{Y},\mathcal{Z}\}$ can be written as

$$\begin{aligned}\hat{\mathbf{X}} &= \frac{1}{2}\left\{cov(\mathbf{x}-\mathbf{y}) + cov(\mathbf{x}-\mathbf{z}) - cov(\mathbf{y}-\mathbf{z})\right\}\\ &\quad + \left[E(\boldsymbol{\epsilon}^x \odot \boldsymbol{\epsilon}^y) + E(\boldsymbol{\epsilon}^x \odot \boldsymbol{\epsilon}^z) - E(\boldsymbol{\epsilon}^y \odot \boldsymbol{\epsilon}^z)\right]\,,\end{aligned} \tag{13a}$$

$$\begin{aligned}\hat{\mathbf{Y}} &= \frac{1}{2}\left\{cov(\mathbf{y}-\mathbf{z}) + cov(\mathbf{y}-\mathbf{x}) - cov(\mathbf{z}-\mathbf{x})\right\}\\ &\quad + \left[E(\boldsymbol{\epsilon}^y \odot \boldsymbol{\epsilon}^z) + E(\boldsymbol{\epsilon}^y \odot \boldsymbol{\epsilon}^x) - E(\boldsymbol{\epsilon}^z \odot \boldsymbol{\epsilon}^x)\right]\,,\end{aligned} \tag{13b}$$

$$\begin{aligned}\hat{\mathbf{Z}} &= \frac{1}{2}\left\{cov(\mathbf{z}-\mathbf{x}) + cov(\mathbf{z}-\mathbf{y}) - cov(\mathbf{x}-\mathbf{y})\right\}\\ &\quad + \left[E(\boldsymbol{\epsilon}^z \odot \boldsymbol{\epsilon}^x) + E(\boldsymbol{\epsilon}^z \odot \boldsymbol{\epsilon}^y) - E(\boldsymbol{\epsilon}^x \odot \boldsymbol{\epsilon}^y)\right]\,.\end{aligned} \tag{13c}$$

With the corners $\{\mathcal{X},\mathcal{Y},\mathcal{Z}\}$ of 3CH defined to be the observations, background, and analysis, $\{\mathcal{O},\mathcal{B},\mathcal{A}\}$, this subsection explores the full relationship between the methods. As in section 2.1, associating the variables $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ with $\mathbf{o}$,

b and **a**, respectively, each of the expressions in (13) become,

$$\hat{\mathbf{X}} \equiv \frac{1}{2}\left[cov(\mathbf{o}-\mathbf{b}) + cov(\mathbf{o}-\mathbf{a}) - cov(\mathbf{b}-\mathbf{a})\right] + \Delta\mathbf{X}, \tag{14a}$$

$$\hat{\mathbf{Y}} \equiv \frac{1}{2}\left[cov(\mathbf{b}-\mathbf{a}) + cov(\mathbf{b}-\mathbf{o}) - cov(\mathbf{a}-\mathbf{o})\right] + \Delta\mathbf{Y}, \tag{14b}$$

$$\hat{\mathbf{Z}} \equiv \frac{1}{2}\left[cov(\mathbf{a}-\mathbf{o}) + cov(\mathbf{a}-\mathbf{b}) - cov(\mathbf{o}-\mathbf{b})\right] + \Delta\mathbf{Z}, \tag{14c}$$

with the explicit form of the cross–covariance terms written as

$$\Delta\mathbf{X} = E\left(\boldsymbol{\epsilon}^o \odot \boldsymbol{\epsilon}^b\right) + E\left(\boldsymbol{\epsilon}^a \odot (\boldsymbol{\epsilon}^o - \boldsymbol{\epsilon}^b)\right), \tag{15a}$$

$$\Delta\mathbf{Y} = E\left(\boldsymbol{\epsilon}^o \odot \boldsymbol{\epsilon}^b\right) - E\left(\boldsymbol{\epsilon}^a \odot (\boldsymbol{\epsilon}^o - \boldsymbol{\epsilon}^b)\right), \tag{15b}$$

$$\Delta\mathbf{Z} = E\left(\boldsymbol{\epsilon}^a \odot (\boldsymbol{\epsilon}^o + \boldsymbol{\epsilon}^b)\right) - E\left(\boldsymbol{\epsilon}^o \odot \boldsymbol{\epsilon}^b\right), \tag{15c}$$

after a convenient rearrangement of terms. Let us now find out what 3CH obtains when the assumptions of DBCP are applied.

The assumption of uncorrelated observation and background errors, immediately leads to a simplification of the cross–covariance terms in (15) to

$$\Delta\mathbf{X} = E\left(\boldsymbol{\epsilon}^a \odot (\boldsymbol{\epsilon}^o - \boldsymbol{\epsilon}^b)\right), \tag{16a}$$

$$\Delta\mathbf{Y} = E\left(\boldsymbol{\epsilon}^a \odot (\boldsymbol{\epsilon}^b - \boldsymbol{\epsilon}^o)\right), \tag{16b}$$

$$\Delta\mathbf{Z} = E\left(\boldsymbol{\epsilon}^a \odot (\boldsymbol{\epsilon}^o + \boldsymbol{\epsilon}^b)\right), \tag{16c}$$

from where it follows that $\Delta\mathbf{Y} = -\Delta\mathbf{X}$. Furthermore, orthogonality between the analysis error and the innovation vector results in

$$\Delta\mathbf{X} = 0, \tag{17a}$$

$$\Delta\mathbf{Y} = 0, \tag{17b}$$

$$\Delta\mathbf{Z} = 2E\left(\boldsymbol{\epsilon}^a \odot \boldsymbol{\epsilon}^b\right), \tag{17c}$$

with the last expression following from recognizing that this orthogonality also implies that $E(\boldsymbol{\epsilon}^a \odot \boldsymbol{\epsilon}^o) = E(\boldsymbol{\epsilon}^a \odot \boldsymbol{\epsilon}^b)$.

Substituting (8) and using the expression for the Kalman gain (9) the rhs of (17c) becomes

$$\begin{aligned} \Delta\mathbf{Z} &= 2E\left(\boldsymbol{\epsilon}^a \odot \boldsymbol{\epsilon}^b\right) \\ &= (\mathbf{I}-\mathbf{K})\mathbf{B} + \mathbf{B}(\mathbf{I}-\mathbf{K})^T \\ &= 2\mathbf{A}, \end{aligned} \tag{18}$$

which now fully determines what the cross–covariance error terms in 3CH amount to under the assumptions of DBCP. Here we see that, this somewhat unnatural choice of corners for 3CH turns out to luckily fit its *assumption* that errors be uncorrelated, at least when it comes to its first two corners; the same, however, cannot be said of the cross terms of the third corner.

To completely evaluate the 3CH expressions it is helpful to notice that a little matrix algebra shows that

$$cov(\mathbf{o}-\mathbf{b}) = cov(\boldsymbol{\epsilon}^o - \boldsymbol{\epsilon}^b) = \mathbf{B}+\mathbf{R}\,, \tag{19a}$$

$$cov(\mathbf{o}-\mathbf{a}) = cov(\boldsymbol{\epsilon}^o - \boldsymbol{\epsilon}^a) = \mathbf{R}-\mathbf{A}\,, \tag{19b}$$

$$cov(\mathbf{a}-\mathbf{b}) = cov(\boldsymbol{\epsilon}^a - \boldsymbol{\epsilon}^b) = \mathbf{B}-\mathbf{A}\,. \tag{19c}$$

Combining these with (17) and (18) into (14), it follows that

$$\hat{\mathbf{X}} = \mathbf{R}\,, \tag{20a}$$

$$\hat{\mathbf{Y}} = \mathbf{B}\,, \tag{20b}$$

$$\hat{\mathbf{Z}} = -\mathbf{A} + 2\mathbf{A}\,, \tag{20c}$$

where the last expression is left explicitly unfolded to emphasize the fact that the 2$\mathbf{A}$ term comes from the cross–covariance terms in the third corner of 3CH, viz. (17c). This term is what actually is neglected when 3CH assumes the errors in its data sets to be uncorrelated; which is not the case when the corners are made of the $\{\mathcal{O}, \mathcal{B}, \mathcal{A}\}$ data sets. Here we see that the cancellation of the cross–terms in the first two corners [viz, (17a) and (17b)] allows for 3CH to obtain the same results as DBCP for the observation and background error covariances.

An alternative, and perhaps simpler, way to see the subtlety in the relationship between the two methods when it comes to the third corner of 3CH and the estimation of the analysis error covariance is to notice that, the blind assumption of uncorrelated errors made in 3CH implies that

$$cov(\mathbf{o}-\mathbf{a}) \overset{3CH}{=} \mathbf{R}+\mathbf{A}\,, \tag{21}$$

which is clearly not the case, when the analysis (third corner) error is correlated with the observation (first corner) error, viz. (19b), as in typical data assimilation algorithms. It is worth to point out that (21) has been recognized in works deriving residual diagnostics when employing observations not assimilated in the underlying DA system (see Marseille et al. 2016; Ménard and Deshaies-Jacques 2018). Furthermore, relation (19b) has been known since at least the times of Hollingsworth and Lönnberg (1989). The present work simply brings these together in another context.

2.4 | General remarks

Having establish the relationship between a particular application of 3CH with DBCP, and having found 3CH to reproduce DBCP, after a well understood adjustment of its third corner result, should render unnecessary any further comparative comments: when the same data is made available to both methods, and all is consistent, there is nothing else to say.

Still, perhaps for clarity, it might be worthwhile to briefly look into how the known limitations affecting each of these methods compare when viewed from the light of the present work. Sjoberg et al. (2021) lists the following factors limiting the accuracy of 3CH estimates: (i) sample size; (ii) outliers in the relevant data sets; (iii) relative magnitude of cross–covariance (random) errors among data sets; (iv) biases; and (v) unknown cross–covariances. Put in the context of 3CH's relationship with DBCP this is what can be stated:

i Sample size is always a factor when estimating errors from a finite sample; 3CH and DBCP are alike in this respect.

ii Outliers can certainly affect both procedures, but are typically not a factor in DBCP given its residual vectors derive directly from DA algorithms. Such residuals have usually been cleared by multiple levels of quality control, reducing the effect of outliers. In the context when 3CH derives its observation, background and analysis from the residuals used in DBCP, the former is only mildly affected by outliers, just as the latter. Use of alternative 3CH data sets should require as much care to the data as the analysis algorithm gives to its residuals.

iii The relative magnitude of the errors in the data sets used for 3CH and DBCP should not be so much an issue in the particular context here. Errors in the observations, background and analysis tend not to be largely different from each other, in some sense, the assimilation homogenizes the errors; this is more of an issue when 3CH is used in its broader context of using alternative data sets.

iv Though unaccounted biases can be an issue in general for both methods, data assimilation residuals benefit from various levels of bias correction: (i) applied to the observations either offline (e.g., Haimberger 2007) or online as in variational procedures (e.g, Dee 2005, and references therein); and (ii) applied to the background (underlying model) as in weak–constraint variational applications (e.g., Bonavita 2021, and references therein). DBCP benefits from these automatically and, as long as 3CH construct its data sets from the same residuals used in DBCP, the effect should be similar in 3CH.

v Unknown cross–covariances in the context of the present work would be a manifestation of lack of optimality in the underlying DA. This in turn could be a consequence of numerous issues, e.g., non-whiteness of residuals, non–linearities, unaccounted errors in the forward model. The relationship between 3CH and DBCP here has been established under the assumption of linearity; this is basic to DBCP. When linearity breaks down, the relationship established here also breaks down.

Additionally, DBCP calculations are typically based on residual statistics obtained from deterministic, high resolution, DA systems. In such cases, only a single realization of residuals is available and the ergodic assumption must be relied upon as time averages are used for what should be expectation. As investigated in Desroziers et al. (2009), ensemble-based systems have the potential to ameliorate this situation by using averages with respect to the ensemble of residuals – though depending on the number of ensemble members, time averaged might still be needed to gather a robust sample size. To the extent that the 3CH version in the present work construct its three data sets from the residuals used in DBCP, the concerns in regards to realizations applies to 3CH just as well. Even in 3CH general form, with arbitrary two corners, single–realization estimates must be taken with caution.

Furthermore, a point of interest when it comes to 3CH is that it makes no assumptions about the nature of the underlying statistics. Interestingly, DBCP's dependence on Gaussianity is quite loose. Derivations of DBCP state that the underlying observation and model errors are Gaussian as a general statement when posing the arguments made in traditional DA techniques. However, as seen in section 2.2 and in the appendix, there are no explicit assumptions on the underlying statistics of errors for the derivation of DBCP. In practice, non–linearity (a version of non–Gaussianity) is handled with multiple outer loops and other ways that are not accounted for in the statements of DBCP. The effect of non–Gaussianity might be something to explore in how it affects DBCP and in how it compares with 3CH.

Most applications of DBCP are intended to tune the prescribed error statistics. In working to improve the prescription of observation error statistics, DBCP has been used to derive not only variance information but also error correlations (off–diagonal). This has become central to recent developments that expand the capabilities of DA and allow for representation of, for example, existing inter–channel correlations present in satellite radiance observations (e.g., Stewart et al. 2014, Weston et al. 2014, Bormann et al. 2016, Campbell et al. 2017, Geer 2019). As highlighted in Semane et al. (2022), the caveats of using DBCP estimates to refine only observation uncertainties while ignoring estimates of background uncertainties are discussed in the works of Ménard (2016), Waller et al. (2016), and Bathmann (2018). The consequences of tuning *only* a subset of the error covariances in DA has also been investigated in Bowler (2017) in the context of Todling (2015b) and Todling (2015a) extension of DBCP's approach to estimate model error covariance in weak constraint DA.

It might be desirable to corroborate error observation estimates from either DBCP or 3CH in alternative ways by, for example, making use of independent observations (e.g., Ménard and Deshaies-Jacques (2018), Mirza et al. (2021). The generality of two of the corners of 3CH offers yet another possibility for a methodology to obtain alternative estimates. This is the gist of the work of Anthes et al. (2022). The question of how to make use of such alternative estimates to aid the prescription of uncertainties assigned to observations used in DA algorithms is left for future investigation.

As a final remark we point out that the concept of *truth* implicit in the calculations in section 2.3 and the appendix, is the academic one employed in estimation theory textbooks (e.g., Jazwinski 1970; Maybeck 1979). No attempt has been made to account for representativeness errors along the lines of, for example, Janjić and Cohn (2006) and its consequences to the expression for the innovation covariance — see eq. (26) in Janjić et al. (2018). A non–textbook concept of truth would help account for errors in how data are sampled, collocated in space and time, and adjusted for footprint representation (e.g., Table 1 in Semane et al. 2022). A similar statement can be made about representing errors in the observation operator along the lines of Waller et al. (2014). It should be possible to combine all these with the suboptimality arguments in the appendix to bring forth more general statements. None of these is likely to result in differences between DBCP and 3CH as long the corners of the latter are made consistent with the information used by the former.

3 | PRACTICAL COMPARISON BETWEEN 3CH AND DBCP

The previous section has established the relationship between DBCP and 3CH when the latter uses a particular choice of corners, and shows the methods to be identical to within a well understood sign difference in one of the estimates. With that, it would seem unnecessary to show numerical illustrations from practical applications for after all, any noticeable differences would seem to indicate either errors in the supporting software, or inconsistency in how data are sampled to produce corresponding results. Nonetheless, in light of the results in Semane et al. (2022) showing small differences between the two methods (see their Figs. 2 and 3), at least a figure corresponding to a revision of their results seems appropriate. This section provides a revision of the results in that work, corroborating the equivalence of numerical results when both methods use the same data sets.

Semane et al. (2022) produce numerical comparisons of 3CH and DBCP for (standard deviation) uncertainties of COSMIC-2 RO bending angle assimilated in ERA5. The ERA5 reanalysis is produced from ECMWF's Integrated Forecasting System cycle 41r2 (the cycle used for operational forecasting in 2016) with a forecast model grid spacing of 31 km and 137 vertical levels. The ERA5 assimilation implements a 12–hour window 4D-Var with cycles from 0900-2100 UTC and 2100-0900 UTC (in the following day), where the background and the observations falling within a time window are used to specify the 4D analyses within the window (see Hersbach et al. 2020).

The COSMIC-2 observations used for the comparison are the provisional level-2 bending angles provided by the University Corporation for Atmospheric Research (UCAR) COSMIC Data Analysis and Archive Center (CDAAC) available since October 1, 2019. The bending angle observations are provided on 247 vertical levels (von Engeln et al. 2009), matching the levels used by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) Radio Occultation Meteorology Satellite Application Facility (ROM SAF) processing of the GNSS Receiver for Atmospheric Sounding (GRAS) receiver onboard the Meteorological Operational (Metop) Satellites. All COSMIC-2

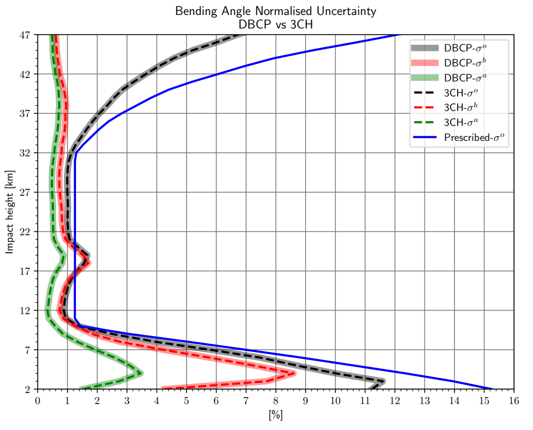


FIGURE 1 Estimated bending angle standard deviations (uncertainties) from the DBCP (solid–shade) and 3CH (dashed) methods for April 2021. The standard deviations of the COSMIC-2, ERA5 background, and ERA5 analysis bending angles are shown by the black, red and green profiles, respectively. The prescribed uncertainty percentage is shown in blue. These estimates are for all COSMIC-2 latitudes (50°S-50°N).

profiles available from 50°S to 50°N are used in the comparison and all data were normalized at each level by the sample mean of the ERA5 background data and averaged over 1 km layers. Only data passing both a "first guess check" on the observed minus bending angles calculated from the short-range forecast and variational quality control during the 4D-Var (Anderson and Järvinen 1999; Ruston and Healy 2020) are used to gather relevant statistics. Figure 1 shows the vertical profiles of the DBCP–estimated uncertainties of the COSMIC-2 (black solid), ERA5 background (red solid), ERA5 analysis (green solid), with the corresponding 3CH estimates (dashed curves), and the prescribed observation uncertainty (blue curve). The latter is a global "model", which only includes variation in the vertical as function of impact height. We see that, from the surface to 10 km, the prescribed uncertainty decreases linearly with impact height to 1.25%; above 10 km, 1.25% is used until this reaches the 3 microradian lower limit. The various uncertainty estimates show a maximum in the lower troposphere around 4 km impact height (~2 km mean sea level height), which is associated with temperature and moisture variability at the top of the atmospheric boundary layer. A small relative maximum appears around 18 km impact height. This level is in the vicinity of the tropopause where temperature is highly variable.

Three-cornered hat estimates can be derived by extracting the observations, background and analysis from the residuals used to produce DBCP estimates, and associate those with the corners of 3CH. The 3CH (dashed curves) estimates in Fig. 1 lay right over those of DBCP; the 3CH analysis curve is produced after taking the square root of the *negative* of the third corner variance estimate. Results here show that, when both DBCP and 3CH rely on identical data sets to produce their estimates, results between the methods corroborate the mathematical derivation in Section 2. The small differences found in Semane et al. (2022), when comparing observation and background errors, are now understood to be due to sampling differences in the implementation of the two methods back then; an estimate of analysis error from 3CH had not been produced back then since the negative result was not understood at the time.

The results in Fig. 1 only compare the (square–root of the) diagonal of the error covariances of DBCP and 3CH, but the relationship between the methods holds for the whole covariances. As an additional illustration, not studied in Semane et al. (2022), Fig. 2a shows DBCP's observation uncertainty in vertical correlations of bending angle. The

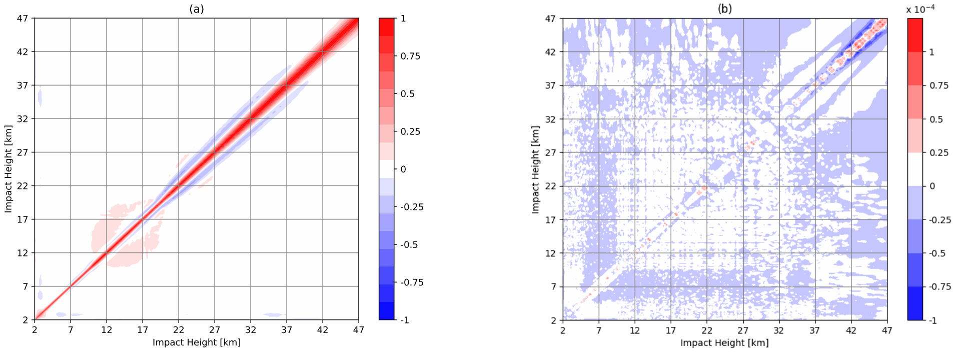


FIGURE 2 Left: Vertical correlations in COSMIC-2 bending angle observations assimilated in ERA5, in April 2021, estimated using DBCP method. Right: difference of estimated correlations between 3CH and DBCP (notice 10^{-4} scale in color bar).

matrix is seen to be nearly diagonal with a minor correlation increase in adjacent levels when going from 2 km up to 47 km; between 20 and 38 km there a slight anti–correlation among levels slightly further apart. Between 12 to about 18 km there is also some small (~ 0.1) correlation within a few nearby levels. All correlations are weak and seem to corroborate the present diagonal observation error covariance prescription used when assimilating bending angle in ERA5, and elsewhere. The corresponding 3CH estimate produces nearly identical results, so much so that only the difference between the two estimates is shown in Fig. 2b. A rather tight shading interval of 10^{-4} reveals a minor difference along the diagonal toward the top of the grid. These differences are seen to be due to how errors accumulate in the coding of the two methods. It is relevant to point out that Nielsen et al. (2022) have recently used 3CH to calculate similar RO correlations but for refractivities. Since refractivity is a weighted sum of bending angles, there is more correlations in refractivity space than seen here in bending angle space.

Just as independent corroboration, and to fulfill curiosity, the 3CH method has been (optionally) added to the suite of programs that systematically produce DBCP results for residuals generated by the NASA GEOS 4D Hybrid Ensemeble-Variational system (Todling et al. 1998). In this implementation the data needed by both methods is identically sampled. Any differences found between the two methods in either standard deviation or correlations of any type are at round-off levels: only due to how statistics are accumulated in calculating the terms of DBCP versus those of 3CH. Showing any of such results is deemed unnecessary.

4 | CLOSING REMARKS

The present work shows that, when the three–cornered hat (3CH) method of Gray and Allan (1974) uses the observation, background and analysis for its three corners, it recovers identical uncertainty estimates for observations and backgrounds as obtained with the method of (Desroziers et al., 2005, DBCP) under similar assumptions. The third corner estimate recovers instead the negative of the analysis error covariance. These surprising results occur because in the 3CH method, the neglected error cross-covariance between observations and analysis and between background and analysis, which are both positive, cancel in the 3CH equations for the estimates of observations and background uncertainties, but add in the equation for the estimates of the analysis uncertainty. In this latter case, their neglect by 3CH amounts to its result becoming the negative of the DBCP uncertainty estimate for the analysis error covariance.

The relationship established here means that the role of 3CH when it comes to evaluating errors of interest to data assimilation procedures is not of replacing DBCP, but rather to allow for alternative means of producing such estimates. The freedom in choosing two of the corners of 3CH makes it attractive in attempts to find alternative ways to corroborate the estimates from DBCP.

Acknowledgements

Thanks are due to Jeremiah Sjoberg who provided valuable assistance and insights with the 3CH calculations in the ECMWF application. Thanks are also due to three reviewers whose valuable comments helped us clarify various points. Todling was supported by the NASA GMAO Core funding and is thankful to the High Performance Computing facilities provide by the NASA Center for Climate Simulation. Anthes was supported by NSF Grant AGS-2054356 and NOAA Contract 16CN0070. Semane was supported by the Radio Occultation Meteorology Satellite Application Facility (ROM SAF), which is a decentralized processing center under the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT).

A | CONSEQUENCES OF SUBOPTIMALITY

Under the general setting of suboptimality, the statements in section 2.2 must be presented differently. The assumptions of unbiased and uncorrelated observation and background errors are still taken to hold. The linear expression relating these errors can, however, be written more generally as in

$$\boldsymbol{\epsilon}^a = \boldsymbol{\epsilon}^b + \tilde{\mathbf{K}}(\boldsymbol{\epsilon}^o - \boldsymbol{\epsilon}^b)\,, \tag{A.1}$$

where $\tilde{\mathbf{K}}$ is now a general weighting matrix defined as

$$\tilde{\mathbf{K}} \equiv \tilde{\mathbf{B}}\tilde{\boldsymbol{\Gamma}}^{-1}\,, \tag{A.2}$$

for $\tilde{\mathbf{B}}$ and $\tilde{\boldsymbol{\Gamma}}$ being symmetric positive–definite matrices, but otherwise unspecified. Typically, $\tilde{\boldsymbol{\Gamma}} = \tilde{\mathbf{B}} + \tilde{\mathbf{R}}$, where $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{R}}$ are *prescribed* weighting matrices representing uncertainties in the background and observations, respectively, but not necessarily corresponding to the *true* error covariances. Optimality means that $\tilde{\mathbf{K}} = \mathbf{K}$, which is a statement that can be broken up in two:

OPT-1: That the *innovation covariance consistency* (icc) statement (Ménard 2016) should hold, that is,

$$\tilde{\boldsymbol{\Gamma}} \overset{icc}{=} \boldsymbol{\Gamma} \equiv E[(\boldsymbol{\epsilon}^o - \boldsymbol{\epsilon}^b)(\boldsymbol{\epsilon}^o - \boldsymbol{\epsilon}^b)^T]\,. \tag{A.3}$$

where $\boldsymbol{\Gamma} = \mathbf{B} + \mathbf{R}$ is the innovation error covariance.

OPT-2: And that the observation–space projected background error covariance (opb) should equal the corresponding projection of the true background error covariance, that is,

$$\tilde{\mathbf{B}} \overset{opb}{=} \mathbf{B} \,. \tag{A.4}$$

The split of optimality in these two subcategories allows for thinking of systems that are well tuned (icc) even when not fully optimal, i.e., even when opb is not satisfied.

With the above, a piecemeal derivation of (12) reveals:

$$\begin{aligned}
\hat{\mathbf{R}} &= \frac{1}{2}\left[(\mathbf{I}-\tilde{\mathbf{K}})\boldsymbol{\Gamma} + \boldsymbol{\Gamma}(\mathbf{I}-\tilde{\mathbf{K}})^T\right] \\
&\overset{icc}{=} \tilde{\mathbf{R}} \,, & \text{(A.5a)} \\
\hat{\mathbf{B}} &= \frac{1}{2}\left[\tilde{\mathbf{K}}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}\tilde{\mathbf{K}}^T\right] \\
&\overset{icc}{=} \tilde{\mathbf{B}} \,, & \text{(A.5b)} \\
\hat{\mathbf{A}} &= \frac{1}{2}\left[\tilde{\mathbf{K}}\boldsymbol{\Gamma}(\mathbf{I}-\tilde{\mathbf{K}})^T + (\mathbf{I}-\tilde{\mathbf{K}})\boldsymbol{\Gamma}\tilde{\mathbf{K}}^T\right] \\
&\overset{icc}{=} \frac{1}{2}\left[\tilde{\mathbf{B}}(\mathbf{I}-\tilde{\mathbf{B}}\tilde{\boldsymbol{\Gamma}}^{-1})^T + (\mathbf{I}-\tilde{\mathbf{B}}\tilde{\boldsymbol{\Gamma}}^{-1})\tilde{\mathbf{B}}\right] \\
&\overset{icc}{=} \tilde{\mathbf{A}} \,, & \text{(A.5c)}
\end{aligned}$$

where $\tilde{\mathbf{A}} \equiv (\mathbf{I}-\tilde{\mathbf{K}})\tilde{\mathbf{B}}$ is the *perceived* analysis error covariance, i.e., the error the DA "thinks" it is making. By replacing $\boldsymbol{\Gamma}$ with $\tilde{\boldsymbol{\Gamma}}$, we see that the results above *only* need under the icc statement; full optimality — opb — is not required.

To examine 3CH in the same light of suboptimality, it is helpful to make use of the following relations:

$$E\left(\boldsymbol{\epsilon}^{a} \odot \boldsymbol{\epsilon}^{o}\right) = \frac{1}{2}(\tilde{\mathbf{K}}\mathbf{R} + \mathbf{R}\tilde{\mathbf{K}}^T) \,, \tag{A.6a}$$

$$E\left(\boldsymbol{\epsilon}^{a} \odot \boldsymbol{\epsilon}^{b}\right) = \mathbf{B} - \frac{1}{2}(\tilde{\mathbf{K}}\mathbf{B} + \mathbf{B}\tilde{\mathbf{K}}^T) \,, \tag{A.6b}$$

and

$$cov(\mathbf{o}-\mathbf{b}) = cov(\boldsymbol{\epsilon}^{o} - \boldsymbol{\epsilon}^{b}) = \boldsymbol{\Gamma} \,, \tag{A.7a}$$

$$cov(\mathbf{o}-\mathbf{a}) = cov(\boldsymbol{\epsilon}^{o} - \boldsymbol{\epsilon}^{a}) = (\mathbf{I}-\tilde{\mathbf{K}})\boldsymbol{\Gamma}(\mathbf{I}-\tilde{\mathbf{K}})^T \,, \tag{A.7b}$$

$$cov(\mathbf{a}-\mathbf{b}) = cov(\boldsymbol{\epsilon}^{a} - \boldsymbol{\epsilon}^{b}) = \tilde{\mathbf{K}}\boldsymbol{\Gamma}\tilde{\mathbf{K}}^T \,. \tag{A.7c}$$

Expressions (A.6) and (A.7) are general in that they involve no statements associated with optimality.

Combining (A.6) and (A.7) with (14) and (17), 3CH recovers,

$$
\begin{aligned}
\hat{\mathbf{X}} &= \left\{\frac{1}{2}\left[2\boldsymbol{\Gamma} - \tilde{\mathbf{K}}\boldsymbol{\Gamma} - \boldsymbol{\Gamma}\tilde{\mathbf{K}}^T\right]\right\} + \left\{\frac{1}{2}\left[\tilde{\mathbf{K}}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}\tilde{\mathbf{K}}^T\right] - \mathbf{B}\right\} \\
&\overset{icc}{=} \{\tilde{\mathbf{R}}\} + \{\tilde{\mathbf{B}} - \mathbf{B}\} \\
&\overset{opb}{=} \mathbf{R}\,, && \text{(A.8a)} \\
\hat{\mathbf{Y}} &= \left\{\frac{1}{2}\left[\tilde{\mathbf{K}}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}\tilde{\mathbf{K}}^T\right]\right\} + \left\{\mathbf{B} - \frac{1}{2}\left[\tilde{\mathbf{K}}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}\tilde{\mathbf{K}}^T\right]\right\} \\
&\overset{icc}{=} \{\tilde{\mathbf{B}}\} + \{\mathbf{B} - \tilde{\mathbf{B}}\} \\
&\overset{opb}{=} \mathbf{B}\,, && \text{(A.8b)} \\
\hat{\mathbf{Z}} &= \left\{\frac{1}{2}\left[(\mathbf{I} - \tilde{\mathbf{K}})\boldsymbol{\Gamma}(\mathbf{I} - \tilde{\mathbf{K}})^T + \tilde{\mathbf{K}}\boldsymbol{\Gamma}\tilde{\mathbf{K}}^T - \boldsymbol{\Gamma}\right]\right\} + \left\{\frac{1}{2}\left[(\tilde{\mathbf{K}}\mathbf{R} + \mathbf{R}\tilde{\mathbf{K}}^T) - (\tilde{\mathbf{K}}\mathbf{B} + \mathbf{B}\tilde{\mathbf{K}}^T)\right] + \mathbf{B}\right\} \\
&\overset{icc}{=} \{-\tilde{\mathbf{A}}\} + \{\tilde{\mathbf{A}} + \tilde{\mathbf{K}}\Delta\mathbf{B} + \mathbf{A} - \mathbf{K}\Delta\mathbf{B}\} \\
&\overset{opb}{=} \mathbf{A}\,, && \text{(A.8c)}
\end{aligned}
$$

where the curly brackets separate the contribution of the cross–covariance terms from the terms that can be calculated in practice. Notice we introduce $\Delta\mathbf{B} \equiv (\tilde{\mathbf{B}} - \mathbf{B})$, and arrange the cross term for the last corner expression in a way that highlights what cancels out when opb holds, that is, when $\Delta\mathbf{B} \overset{opb}{=} 0$. Focusing on the second equality of each of the expressions above, we see what icc leads to when 3CH drops the cross–covariance terms (second curly bracket): the first two corners obtain the *prescribed* observation and background error covariances, and the third corner obtains the negative of the *perceived* analysis error covariance. This is similar to what happens in the optimal case, though the derived covariances have different meaning.

It is amusing to notice that allowing for the terms in the curly brackets in the first equality of (A.8c) to add up, the expression for $\hat{\mathbf{Z}}$ can be put in the form

$$
\begin{aligned}
\hat{\mathbf{Z}} &= \tilde{\mathbf{K}}\boldsymbol{\Gamma}\tilde{\mathbf{K}}^T - \tilde{\mathbf{K}}\mathbf{B} - \mathbf{B}\tilde{\mathbf{K}}^T + \mathbf{B} \\
&= (\mathbf{I} - \tilde{\mathbf{K}})\mathbf{B}(\mathbf{I} - \tilde{\mathbf{K}})^T + \tilde{\mathbf{K}}\mathbf{R}\tilde{\mathbf{K}}^T\,, && \text{(A.9)}
\end{aligned}
$$

which is recognized as the Joseph formula encountered in linear filtering and smoother studies on performance evaluation due to misspecification of the Kalman gain. This expression for $\hat{\mathbf{Z}}$ states that under *suboptimal* specification of the gain matrix, the third corner of 3CH captures the so–called *actual* analysis error — see Maybeck (1979, Sec. 6.8) for a discussion on performance analysis. This is considerably different from anything derived from DBCP — not that this is their objective. Unfortunately, this is not a result with practical consequences since the cross–covariance terms of 3CH are inaccessible.

The relationship between the methods hold regardless of optimality. In general, just like DBCP, it is the *prescribed* and *perceived* error covariances that are recovered; under optimality these become the corresponding *true* error covariances.

References

Aabaribaoune, M. E., Emili, E. and Guidard, V. (2021) Estimation of the error covariance matrix for IASI radiances and its impact on the assimilation of ozone in a chemistry transport model. *Atmospheric Measurement Techniques*, **14**, 2841–2856. URL:

https://doi.org/10.5194/amt-14-2841-2021.

Anderson, E. and Järvinen, H. (1999) Variational Quality Control. *Quarterly Journal of the Royal Meteorological Society*, **125**, 697–722. URL: https://doi.org/10.1002/qj.49712555416.

Anthes, R. and Rieckh, T. (2018) Estimating observation and model error variances using multiple data sets. *Atmospheric Measurement Techniques*, **11**, 4239–4260. URL: https://doi.org/10.5194/amt-11-4239-2018.

Anthes, R., Sjoberg, J., Feng, X. and Syndergaard, S. (2022) Comparison of COSMIC and COSMIC-2 Radio Occultation Refractivity and Bending Angle Uncertainties in August 2006 and 2021. *Atmosphere*, **13**, 790. URL: https://doi.org/10.3390/atmos13050790.

Bathmann, K. (2018) Justification for estimating observation-error covariances with the Desroziers diagnostic. *Quarterly Journal of the Royal Meteorological Society*, **144**, 1965–1974. URL: https://doi.org/10.1002/qj.3395.

Bonavita, M. (2021) Exploring the structure of time-correlated model errors in the ECMWF data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **147**, 3454–3471. URL: https://doi.org/10.1002/qj.4137.

Bormann, N., Bonavita, M., Dragani, R., Eresmaa, R., Matricardi, M. and McNally, A. (2016) Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, **142**, 1767–1780. URL: https://doi.org/10.1002/qj.2774.

Bowler, N. E. (2017) On the diagnosis of model error statistics using weak-constraint data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143**, 1916–1928. URL: https://doi.org/10.1002/qj.3051.

Campbell, W. F., Satterfield, E. A., Ruston, B. and Baker, N. L. (2017) Accounting for Correlated Observation Error in a Dual-Formulation 4D Variational Data Assimilation System. *Monthly Weather Review*, **145**, 1019–1032. URL: https://doi.org/10.1175/mwr-d-16-0240.1.

Cheng, S. and Qiu, M. (2021) Observation error covariance specification in dynamical systems for data assimilation using recurrent neural networks. *Neural Computing and Applications*. URL: https://doi.org/10.1007/s00521-021-06739-4.

Dee, D. P. (2005) Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3323–3343. URL: https://doi.org/10.1256/qj.05.137.

Desroziers, G., Berre, L., Chabot, V. and Chapnik, B. (2009) A Posteriori Diagnostics in an Ensemble of Perturbed Analyses. *Monthly Weather Review*, **137**, 3420–3436. URL: https://doi.org/10.1175/2009mwr2778.1.

Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3385–3396. URL: https://doi.org/10.1256/qj.05.108.

von Engeln, A., Healy, S., Marquardt, C., Andres, Y. and Sancho, F. (2009) Validation of operational GRAS radio occultation data. *Geophysical Research Letters*, **36**. URL: https://doi.org/10.1029/2009gl039968.

Gauthier, P., Du, P., Heilliette, S. and Garand, L. (2018) Convergence Issues in the Estimation of Interchannel Correlated Observation Errors in Infrared Radiance Data. *Monthly Weather Review*, **146**, 3227–3239. URL: https://doi.org/10.1175/mwr-d-17-0273.1.

Geer, A. J. (2019) Correlated observation error models for assimilating all-sky infrared radiances. *Atmospheric Measurement Techniques*, **12**, 3629–3657. URL: https://doi.org/10.5194/amt-12-3629-2019.

Gray, J. and Allan, D. (1974) A Method for Estimating the Frequency Stability of an Individual Oscillator. In *28th Annual Symposium on Frequency Control*, 243–246. IEEE. URL: https://doi.org/10.1109/freq.1974.200027.

Haimberger, L. (2007) Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics. *Journal of Climate*, **20**, 1377–1403. URL: https://doi.org/10.1175/jcli4050.1.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049. URL: `https://doi.org/10.1002/qj.3803`.

Hollingsworth, A. and Lönnberg, P. (1989) The verification of objective analyses: Diagnostics of analysis system performance. *Meteorology and Atmospheric Physics*, **40**, 3–27. URL: `https://doi.org/10.1007/bf01027466`.

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A. and Weston, P. (2018) On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **144**, 1257–1278. URL: `https://doi.org/10.1002/qj.3130`.

Janjić, T. and Cohn, S. E. (2006) Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review*, **134**, 2900–2915. URL: `https://doi.org/10.1175/mwr3229.1`.

Jazwinski, A. (1970) *Stochastic processes and filtering theory*. No. 64 in Mathematics in science and engineering. New York, NY [u.a.]: Acad. Press. URL: `http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+021832242&sourceid=fbw_bibsonomy`.

Kalman, R. E. (1960) A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, **82**, 35–45. URL: `https://doi.org/10.1115/1.3662552`.

Lewis, J. M., Lakshmivarahan, S. and Dhall, S. (2006) *Dynamic Data Assimilation*. Cambridge University Press. URL: `https://doi.org/10.1017/cbo9780511526480`.

Marseille, G.-J., Barkmeijer, J., de Haan, S. and Verkley, W. (2016) Assessment and tuning of data assimilation systems using passive observations. *Quarterly Journal of the Royal Meteorological Society*, **142**, 3001–3014. URL: `https://doi.org/10.1002/qj.2882`.

Maybeck, P. S. (1979) *Stochastic Models, Estimatation and Control*, vol. 1. Academic Press.

Ménard, R. (2016) Error covariance estimation methods based on analysis residuals: theoretical foundation and convergence properties derived from simplified observation networks. *Quarterly Journal of the Royal Meteorological Society*, **142**, 257–273. URL: `https://doi.org/10.1002/qj.2650`.

Ménard, R. and Deshaies-Jacques, M. (2018) Evaluation of Analysis by Cross-Validation, Part II: Diagnostic and Optimization of Analysis Error Covariance. *Atmosphere*, **9**, 70. URL: `https://doi.org/10.3390/atmos9020070`.

Mirza, A. K., Dance, S. L., Rooney, G. G., Simonin, D., Stone, E. K. and Waller, J. A. (2021) Comparing diagnosed observation uncertainties with independent estimates: A case study using aircraft-based observations and a convection-permitting data assimilation system. *Atmospheric Science Letters*, **22**. URL: `https://doi.org/10.1002/asl.1029`.

Nielsen, J. K., Gleisner, H., Syndergaard, S. and Lauritsen, K. B. (2022) Estimation of refractivity uncertainties and vertical error correlations in collocated radio occultations, radiosondes and model forecasts. URL: `https://doi.org/10.5194/amt-2022-121`.

Ruston, B. and Healy, S. (2020) Forecast Impact of FORMOSAT-7/COSMIC-2 GNSS Radio Occultation Measurements. *Atmospheric Science Letters*, **22**. URL: `https://doi.org/10.1002/asl.1019`.

Semane, N., Anthes, R., Sjoberg, J., Healy, S. and Ruston, B. (2022) Comparison of Desroziers and Three-Cornered Hat Methods for Estimating COSMIC-2 Bending Angle Uncertainties. *Journal of Atmospheric and Oceanic Technology*. URL: `https://doi.org/10.1175/JTECH-d-21-0175.1`.

Sjoberg, J. P., Anthes, R. A. and Rieckh, T. (2021) The Three-Cornered Hat Method for Estimating Error Variances of Three or More Atmospheric Datasets. Part I: Overview and Evaluation. *Journal of Atmospheric and Oceanic Technology*, **38**, 555–572. URL: https://doi.org/10.1175/jtech-d-19-0217.1.

Stewart, L. M., Dance, S. L., Nichols, N. K., Eyre, J. R. and Cameron, J. (2014) Estimating interchannel observation-error correlations for IASI radiance data in the met office system†. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1236–1244. URL: https://doi.org/10.1002/qj.2211.

Stoffelen, A. (1998) Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research: Oceans*, **103**, 7755–7766. URL: https://doi.org/10.1029/97jc03180.

Tabeart, J. M., Dance, S. L., Lawless, A. S., Migliorini, S., Nichols, N. K., Smith, F. and Waller, J. A. (2020a) The impact of using reconditioned correlated observation-error covariance matrices in the met office 1d-var system. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1372–1390. URL: https://doi.org/10.1002/qj.3741.

Tabeart, J. M., Dance, S. L., Lawless, A. S., Nichols, N. K. and Waller, J. A. (2020b) Improving the condition number of estimated covariance matrices. *Tellus A: Dynamic Meteorology and Oceanography*, **72**, 1–19. URL: https://doi.org/10.1080/16000870.2019.1696646.

Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M. and Zhen, Y. (2020) A Review of Innovation-Based Methods to Jointly Estimate Model and Observation Error Covariance Matrices in Ensemble Data Assimilation. *Monthly Weather Review*, **148**, 3973–3994. URL: https://doi.org/10.1175/mwr-d-19-0240.1.

Todling, R. (2015a) A complementary note to 'A lag-1 smoother approach to system-error estimation': The intrinsic limitations of residual diagnostics. *Quarterly Journal of the Royal Meteorological Society*, **141**, 2917–2922. URL: https://doi.org/10.1002/qj.2546.

— (2015b) A lag-1 smoother approach to system-error estimation: sequential method. *Quarterly Journal of the Royal Meteorological Society*, **141**, 1502–1513. URL: https://doi.org/10.1002/qj.2460.

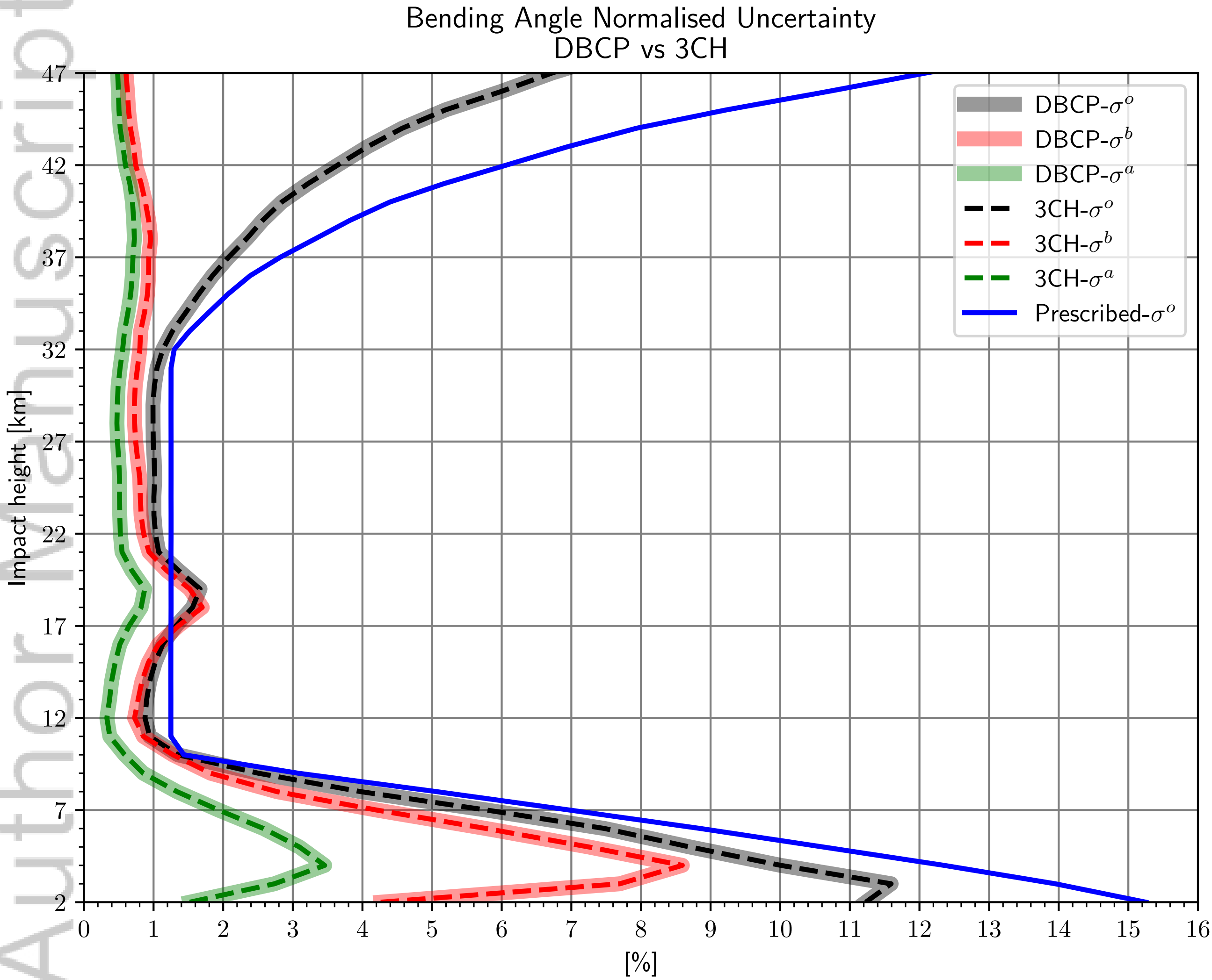
Todling, R., Cohn, S. E. and Sivakumaran, N. S. (1998) Suboptimal Schemes for Retrospective Data Assimilation Based on the Fixed-Lag Kalman Smoother. *Monthly Weather Review*, **126**, 2274–2286. URL: https://doi.org/10.1175/1520-0493(1998)126<2274:ssfrda>2.0.co;2.

Waller, J. A., Bauernschubert, E., Dance, S. L., Nichols, N. K., Potthast, R. and Simonin, D. (2019) Observation Error Statistics for Doppler Radar Radial Wind Superobservations Assimilated into the DWD COSMO-KENDA System. *Monthly Weather Review*, **147**, 3351–3364. URL: https://doi.org/10.1175/mwr-d-19-0104.1.

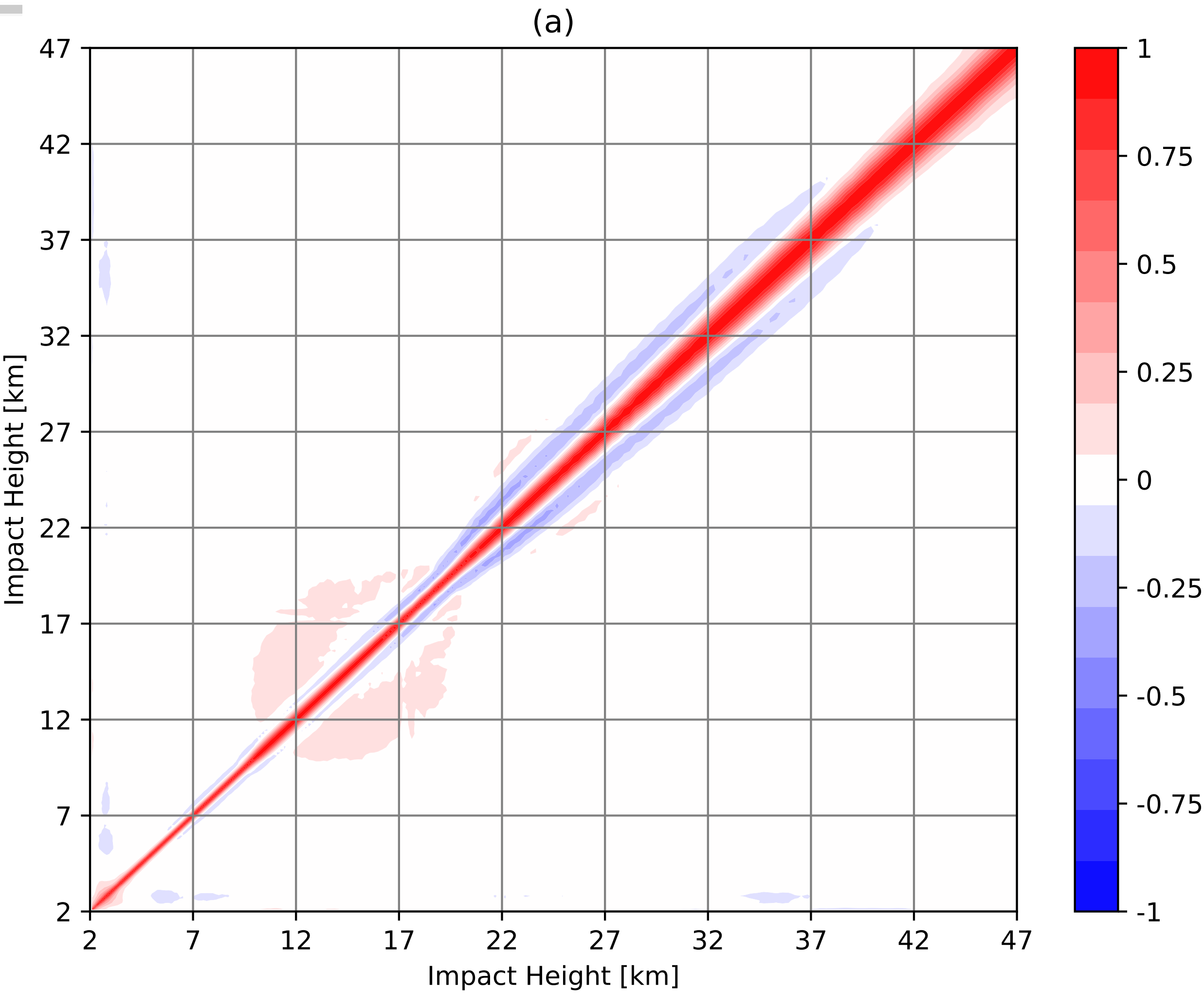
Waller, J. A., Dance, S. L., Lawless, A. S., Nichols, N. K. and Eyre, J. R. (2014) Representativity error for temperature and humidity using the met office high-resolution model†. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1189–1197. URL: https://doi.org/10.1002/qj.2207.

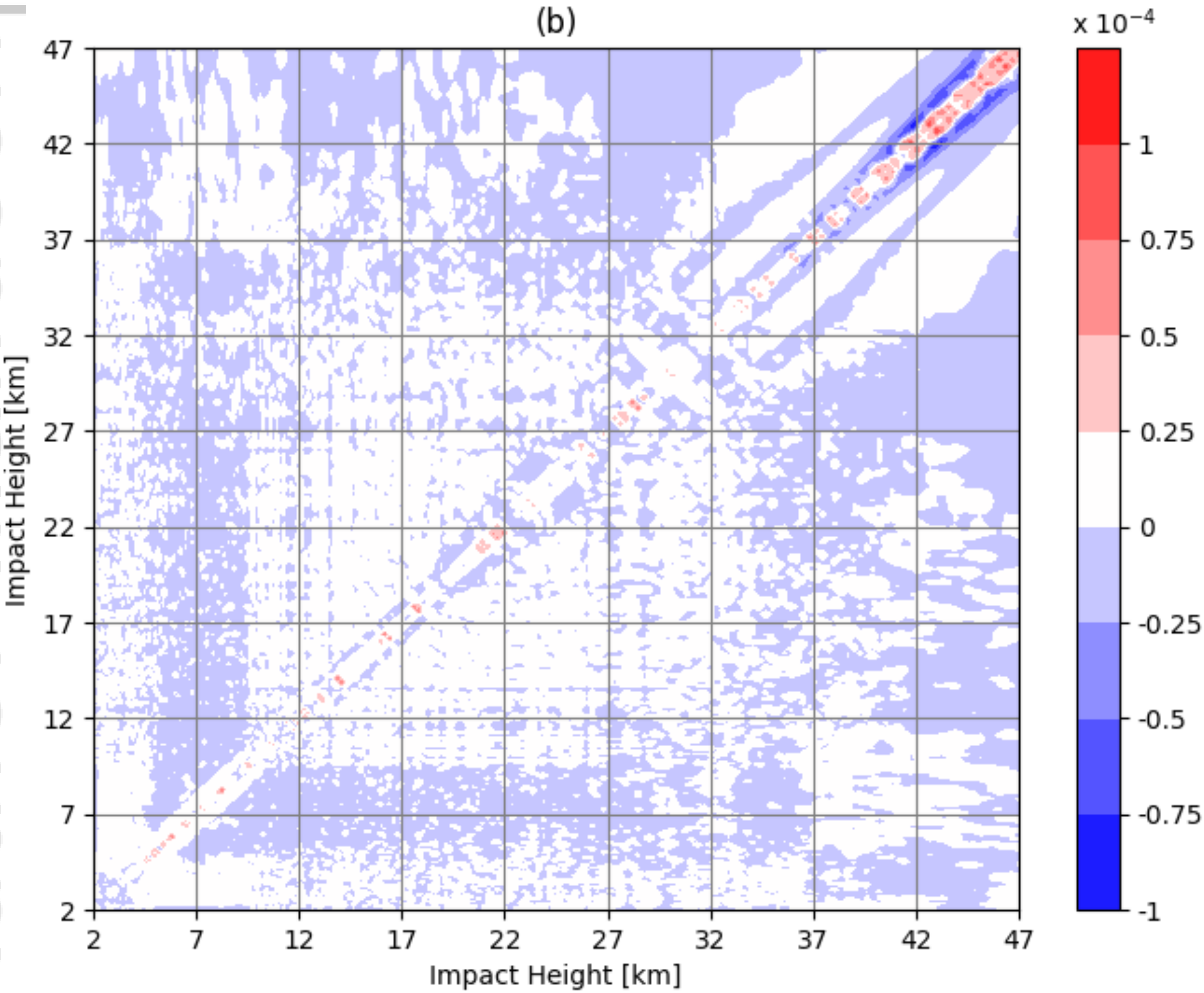
Waller, J. A., Dance, S. L. and Nichols, N. K. (2016) Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Quarterly Journal of the Royal Meteorological Society*, **142**, 418–431. URL: https://doi.org/10.1002/qj.2661.

Weston, P. P., Bell, W. and Eyre, J. R. (2014) Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, **140**, 2420–2429. URL: https://doi.org/10.1002/qj.2306.



Bending Angle Normalised Uncertainty
DBCP vs 3CH
Impact height [km]
[%]
47
42
37
32
27
22
17
12
7
2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
DBCP-σ^o
DBCP-σ^b
DBCP-σ^a
3CH-σ^o
3CH-σ^b
3CH-σ^a
Prescribed-σ^o


(a)
Impact Height [km]
Impact Height [km]
2
7
12
17
22
27
32
37
42
47
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1


(b)
x 10^{-4}
Impact Height [km]
Impact Height [km]
2
7
12
17
22
27
32
37
42
47
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1