Is there a Relationship Between Cornered-hat Methods and a Residual Approach to Estimate System Uncertainty?

Ricardo Todling
Global Modeling and Assimilation Office
NASA/Goddard

9th International Symposium on Data Assimilation (ISDA)
Bologna, Italy
16-20 October 2023

OUTLINE

1. Background
2. Recap: the relationship between 3CH & DBCP
3. Beyond DBPC: Lag-1 smoother residual estimation of system uncertainty
4. Can CH get system uncertainty using information from Lag-1 smoother?
5. Closing Remarks

Background: Cornered-Hat Methods

Atomic Timekeeping and the Statistics of Precision Signal Generators

JAMES A. BARNES

If three oscillators are used, it is possible to independently measure the three quantities σ_{12}, σ_{13}, and σ_{23}. Thus there exist three independent equations:

$$\left.\begin{aligned}\sigma_{12}^2 &= \sigma_1^2 + \sigma_2^2\\ \sigma_{13}^2 &= \sigma_1^2 + \sigma_3^2\\ \sigma_{23}^2 &= \sigma_2^2 + \sigma_3^2\end{aligned}\right\}. \quad (18)$$

A METHOD FOR ESTIMATING THE FREQUENCY STABILITY OF AN INDIVIDUAL OSCILLATOR

James E. Gray and David W. Allan

Time and Frequency Division
National Bureau of Standards
Boulder

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 103, NO. C4, PAGES 7755–7766, APRIL 15, 1998

Toward the true near-surface wind speed: Error modeling and calibration using triple collocation

Ad Stoffelen

Royal Netherlands Meteorological Institute, de Bilt, Netherlands

MARCH 2021 SJOBERG ET AL. 555

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

The idea in so-called Cornered Hat Methods is to use more than one dataset of the same observable to try to estimate the uncertainty in their estimates obtained from them by taking the truth out of the way.

$$\begin{aligned}\mathbf{x} &= \mathbf{t} + \epsilon^{\mathbf{x}}\\ \mathbf{y} &= \mathbf{t} + \epsilon^{\mathbf{y}}\end{aligned}$$

From where it follows:

$$\begin{aligned}cov(\mathbf{t}) = &\frac{1}{4}[cov(\mathbf{x}+\mathbf{y}) - cov(\mathbf{x}-\mathbf{y})]\\ &-\{E[\mathbf{t}\odot(\epsilon^{\mathbf{x}}+\epsilon^{\mathbf{y}})] + E(\epsilon^{\mathbf{x}}\odot\epsilon^{\mathbf{y}})\}\end{aligned}$$

Assuming the datasets have uncorrelated errors, an estimate of the sought uncertainties can be shown to be:

$$\begin{aligned}\hat{\mathbf{X}} &= cov(\mathbf{x}) - \frac{1}{4}[cov(\mathbf{x}+\mathbf{y}) - cov(\mathbf{x}-\mathbf{y})]\\ \hat{\mathbf{Y}} &= cov(\mathbf{y}) - \frac{1}{4}[cov(\mathbf{x}+\mathbf{y}) - cov(\mathbf{x}-\mathbf{y})]\end{aligned}$$

This is the gist of the so-called 2CH – which by neglecting the cross-term between the truth and the errors turns out to have poor accuracy (Sjoberg et al. 2021). Higher order Cornered-Hat Methods only require there be no error correlation among the chosen datasets.

Background: 3CH and DBCP

Semane et al. (2022) compare estimates of observation uncertainty for radio occultation bending angle with the method of Desroziers et al. (2005; DBCP) and the Three-Cornered Hat (3CH) method of Gray & Allan (1974). The Fig. is a comparison.

A back of the envelop calculation during the review process (by the presenter) showed that, when things are ideal, the observation error standard deviation derived with 3CH should equal that derived with DBCP.

A complete analysis of the problem showed that with proper insight 3CH can be taken to recover the results of DBCP.

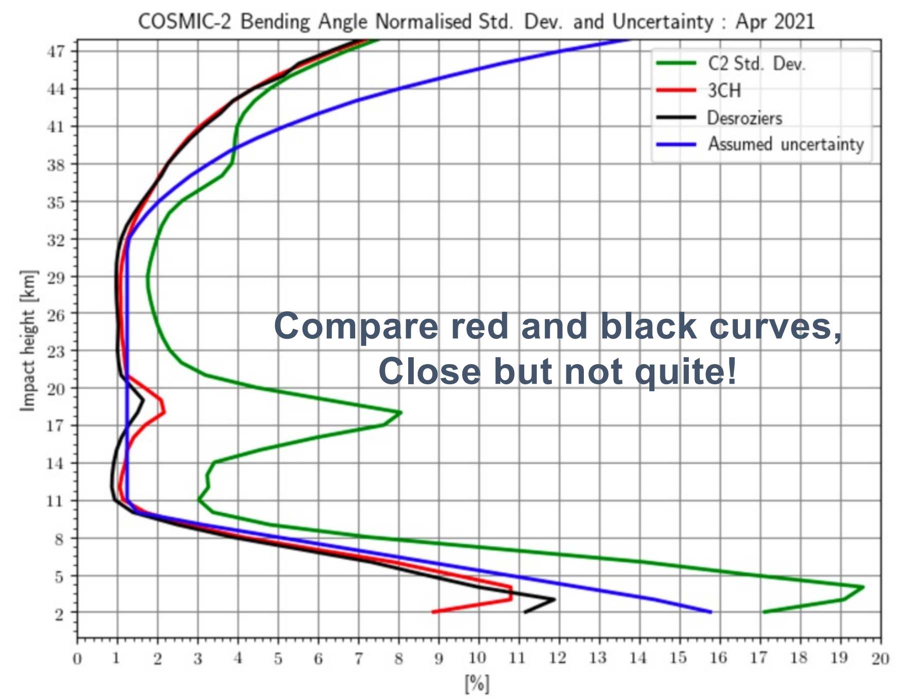


Fig. 2: Estimated COSMIC-2 bending angle random error standard deviations (uncertainties) from the Desroziers (black) and 3CH (red) methods for April 2021. The assumed ECMWF uncertainty model is shown in blue. The standard deviations of the COSMIC-2 bending angles are shown by the green profile. These estimates are for all COSMIC-2 latitudes (50°S-50°N).

Recap: When Corners are Observation, Background and Analysis

Given three datasets $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, the 3CH method uncertainty estimates are given by:

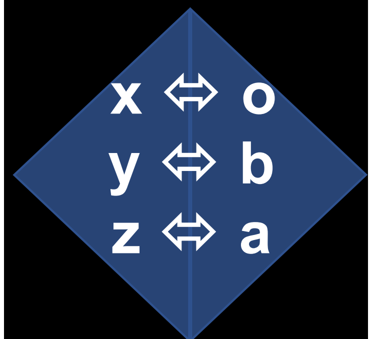
$$
\begin{aligned}
\hat{\mathbf{X}} &= \frac{1}{2}\{cov(\mathbf{x}-\mathbf{y}) + cov(\mathbf{x}-\mathbf{z}) - cov(\mathbf{y}-\mathbf{z})\} \\
&+ \Delta\mathbf{X} \\
\hat{\mathbf{Y}} &= \frac{1}{2}\{cov(\mathbf{y}-\mathbf{z}) + cov(\mathbf{y}-\mathbf{x}) - cov(\mathbf{z}-\mathbf{x})\} \\
&+ \Delta\mathbf{Y} \\
\hat{\mathbf{Z}} &= \frac{1}{2}\{cov(\mathbf{z}-\mathbf{x}) + cov(\mathbf{z}-\mathbf{y}) - cov(\mathbf{x}-\mathbf{y})\} \\
&+ \Delta\mathbf{Z}
\end{aligned}
$$

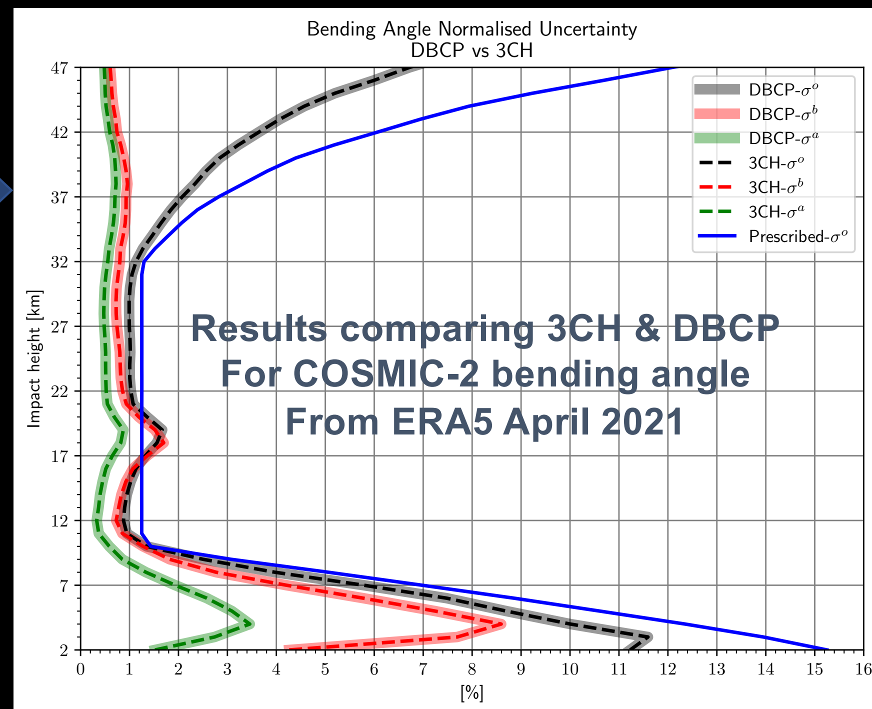
where $cov(\mathbf{u}, \mathbf{v}) = E[(\mathbf{u} - E(\mathbf{u}))(\mathbf{v} - E(\mathbf{v}))^T]$, and with

$$
\begin{aligned}
\Delta\mathbf{X} &= E(\boldsymbol{\epsilon}^{\boldsymbol{x}} \odot \boldsymbol{\epsilon}^{\boldsymbol{y}}) + E(\boldsymbol{\epsilon}^{\boldsymbol{x}} \odot \boldsymbol{\epsilon}^{\boldsymbol{z}}) - E(\boldsymbol{\epsilon}^{\boldsymbol{y}} \odot \boldsymbol{\epsilon}^{\boldsymbol{z}}) \\
\Delta\mathbf{Y} &= E(\boldsymbol{\epsilon}^{\boldsymbol{y}} \odot \boldsymbol{\epsilon}^{\boldsymbol{z}}) + E(\boldsymbol{\epsilon}^{\boldsymbol{y}} \odot \boldsymbol{\epsilon}^{\boldsymbol{x}}) - E(\boldsymbol{\epsilon}^{\boldsymbol{z}} \odot \boldsymbol{\epsilon}^{\boldsymbol{x}}) \\
\Delta\mathbf{Z} &= E(\boldsymbol{\epsilon}^{\boldsymbol{z}} \odot \boldsymbol{\epsilon}^{\boldsymbol{x}}) + E(\boldsymbol{\epsilon}^{\boldsymbol{z}} \odot \boldsymbol{\epsilon}^{\boldsymbol{y}}) - E(\boldsymbol{\epsilon}^{\boldsymbol{x}} \odot \boldsymbol{\epsilon}^{\boldsymbol{y}})
\end{aligned}
$$

being the *unaccessible* random terms.

Practical use of 3CH looks for three datasets with independent errors, so the Δ terms can be safely disregarded.





From Todling et al. (2022).

Reason Why 3CH Almost Gets DBCP

Given that A is correlated with O & B errors, why should {*O*,*B*,*A*} be a viable choice of corners?

Answer: lucky when it comes to first two corners:

- O & B errors are (assumed) uncorrelated.
- Analysis errors are orthogonal to O-B residuals.

However, not so for the third corner:

- Random error add up to twice the analysis error covariance.

$$\begin{aligned}\Delta \mathrm{X} &= E(\epsilon^o \odot \epsilon^b) + E[\epsilon^a \odot (\epsilon^o - \epsilon^b)] \\ \Delta \mathrm{Y} &= E(\epsilon^o \odot \epsilon^b) - E[\epsilon^a \odot (\epsilon^o - \epsilon^b)] \\ \Delta \mathrm{Z} &= E[\epsilon^a \odot (\epsilon^o + \epsilon^b)] - E(\epsilon^o \odot \epsilon^b)\end{aligned}$$

- Uncorrelated O & B errors: $E(\epsilon^o \odot \epsilon^b) = \mathbf{0}$
- Orthogonality: $E[\epsilon^a \odot (\epsilon^o - \epsilon^b)] = \mathbf{0}$

Therefore,

$$\begin{aligned}\Delta \mathrm{X} &= \mathbf{0} \\ \Delta \mathrm{Y} &= \mathbf{0} \\ \Delta \mathrm{Z} &= 2E[\epsilon^a \odot \epsilon^b)] \\ &\overset{icc}{=} 2\tilde{\mathbf{A}} \leftarrow \text{innovation covariance consistency} \\ &\overset{opt}{=} 2\mathbf{A} \leftarrow \text{optimal}\end{aligned}$$

Recap: What does 3CH actually get?

With the association: $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\} \rightarrow \{\mathcal{O}, \mathcal{B}, \mathcal{A}\}$:

Full 3CH

Suboptimal case:

$$
\begin{aligned}
\hat{\mathbf{X}} &= \tilde{\Gamma} - \mathbf{B} \\
\hat{\mathbf{Y}} &= \mathbf{B} \\
\hat{\mathbf{Z}} &= (\mathbf{I} - \tilde{\mathbf{K}})\mathbf{B}(\mathbf{I} - \tilde{\mathbf{K}})^T + \tilde{\mathbf{K}}\mathbf{R}\tilde{\mathbf{K}}^T
\end{aligned}
$$

for $\tilde{\mathbf{K}} = \tilde{\mathbf{B}}\tilde{\Gamma}^{-1}$, i.e., $\hat{\mathbf{Z}}$ arrives at Joseph's formula for the *actual* analysis error covariance (filter performance).

Under Optimality: 3CH = DBCP.

Practical 3CH: neglect of cross–terms

Under Innovation Covariance Consistency:

$$
\begin{aligned}
\hat{\mathbf{X}} &\overset{icc}{=} \tilde{\mathbf{R}} \\
\hat{\mathbf{Y}} &\overset{icc}{=} \tilde{\mathbf{B}} \\
\hat{\mathbf{Z}} &\overset{icc}{=} -\tilde{\mathbf{A}}
\end{aligned}
$$

Under Optimality:

$$
\boxed{
\begin{aligned}
\hat{\mathbf{X}} &\overset{opt}{=} \mathbf{R} \\
\hat{\mathbf{Y}} &\overset{opt}{=} \mathbf{B} \\
\hat{\mathbf{Z}} &\overset{opt}{=} -\mathbf{A}
\end{aligned}
}
$$

Why do these results follow?

In addition to the fact that observation and background errors are assumed uncorrelated …

When it comes to the first two corners of 3CH:

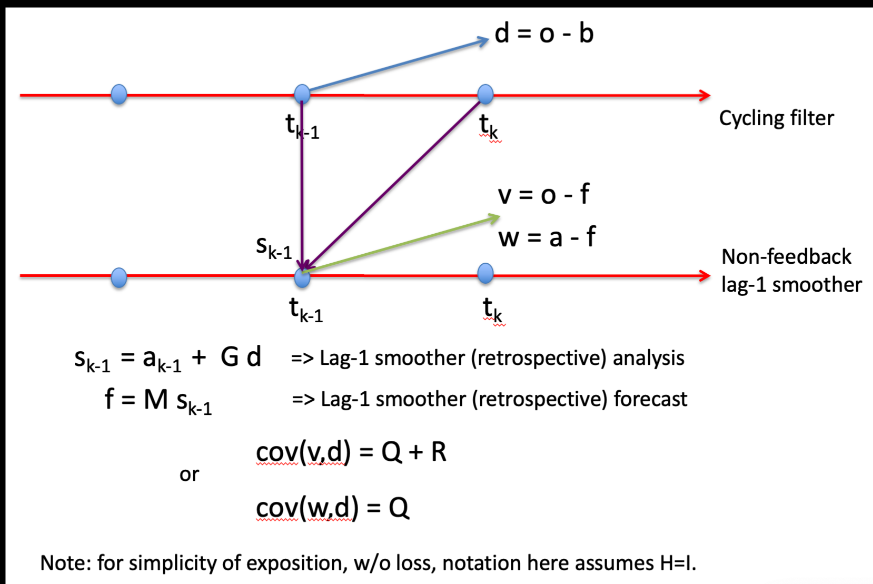
- Random errors cancel out due to the orthogonality between the analysis error and the innovation vector.

When it comes to the third corner of 3CH:

- Gets the negative of the analysis error. With this insight 3CH can be used to recover DBCP.

Beyond DBCP: Estimating System Uncertainty from Lag-1 Smoother

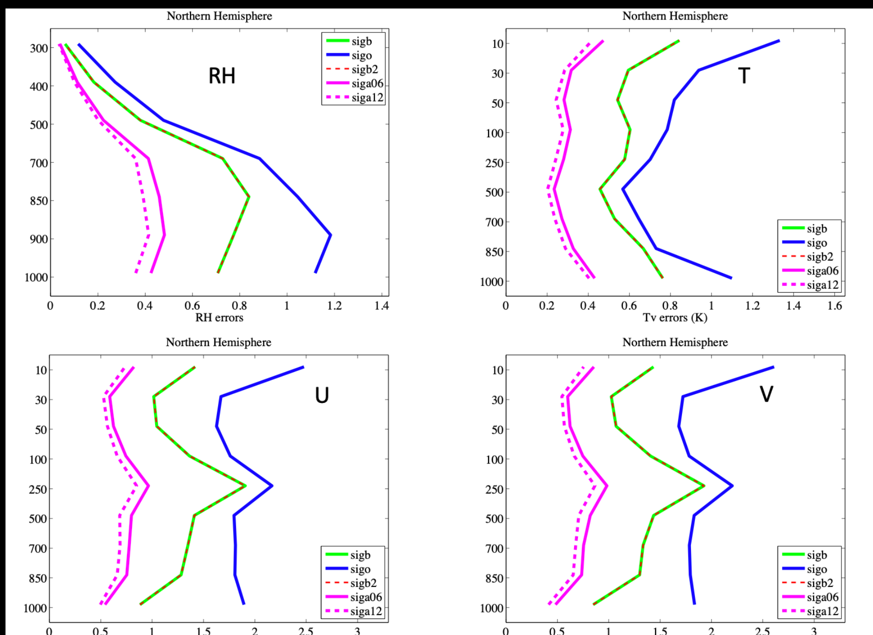
Use of Lag-1 (e.g., 6-hr) Smoother to get Q.



Results from Todling (2015).

- Residuals formed using forecasts f from sequential lag-1 smoother analyses, either v or w, have been shown to provide an estimate of system uncertainty (Q).
- Simple model applications provide illustration for approach.
- Mimic of procedure in IFS (6- and 12-h) 4DVar has provided early estimate of uncertainty standard deviations compared to other DBCP error estimates (work done with Yannick Tremolet ca. 2009).

DBCP from IFS (ca. 2009)

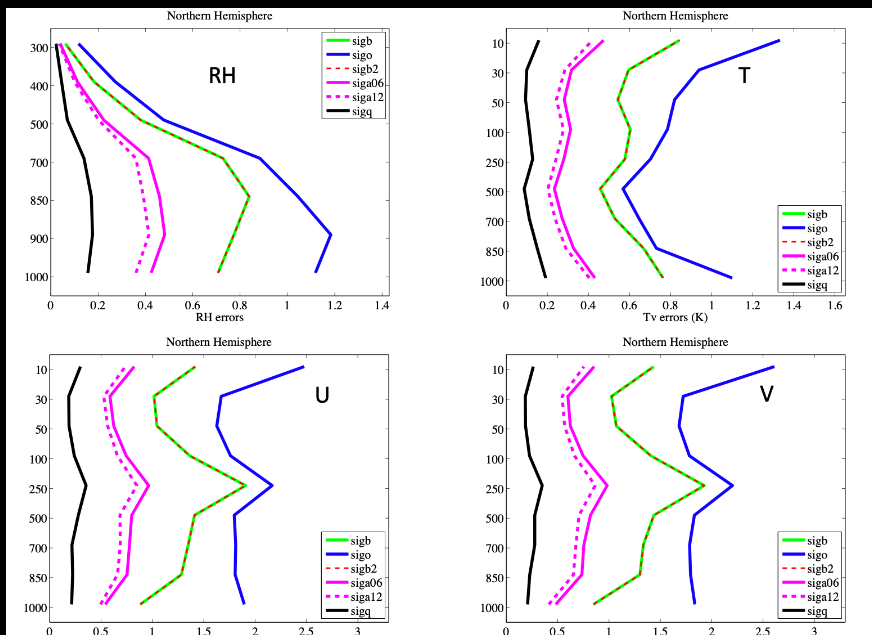


With Yannick Tremolet ca. 2009 (unpublished).

Simulation of Lag-1 approach in IFS

- IFS configured to run a particular configuration of a combined 6- cycling 4DVar with a 12-hour non-cycling 4DVar.
- Examination of residuals from 6-h 4DVar strategies provide DBCP estimates for diag(R), diag(B) and diag(A).
- Examination of residuals from 12-h 4DVar provide estimate of errors from “lag-1” analyses (dashed).
- Results here are for NHE Radiosondes.

Beyond DBCP: Estimates of Q from IFS (ca. 2009)



With Yannick Tremolet ca. 2009 (unpublished).

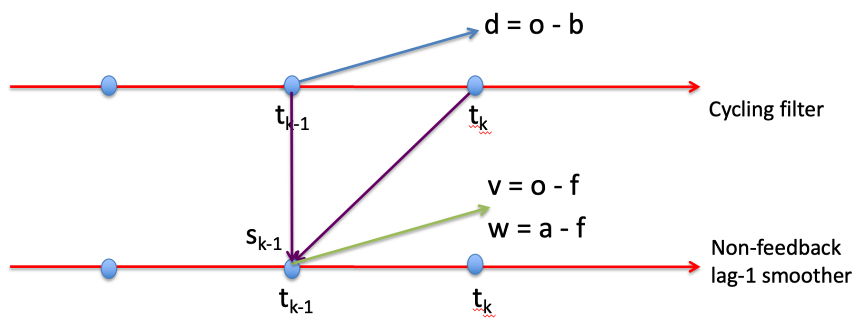
Simulation of Lag-1 approach in IFS

- Further use of residual information from particular implementation of 12-h 4DVar strategy also provides information of diag(**Q**).

- The specific of the 6- and 12-h 4DVars are such that diag(**Q**) should be interpreted as *system uncertainty* (not model error).

- Results here are for NHE Radiosondes.

What should CH consider for possible Q estimate?



$s_{k-1} = a_{k-1} + G\,d$ => Lag-1 smoother (retrospective) analysis

$f = M\,s_{k-1}$ => Lag-1 smoother (retrospective) forecast

$cov(v,d) = Q + R$

or

$cov(w,d) = Q$

Note: for simplicity of exposition, w/o loss, notation here assumes H=I.

Two possible routes

- cov(v,d) involves 3 datasets: o, b, f leading us to think of 3CH as possible candidate to estimate Q with CH method.
- cov(w,d) involves 4 datasets: o, b, a, f leading us to think of 4CH as possible candidate to estimate Q with CH method.
- But CH methods are about getting error (co)variances of the datasets in question.
- That it is, with 3CH we'd at best get R, B, F and with 4CH we'd at best get the same plus A.

About the error covariance CH would get …

If a CH method is to get $\mathbf{F}$, the retrospective forecast error covariance, it would get

$$\mathbf{F} \overset{opt}{=} \mathbf{A} + \mathbf{Q}\Gamma^{-1}\mathbf{Q}$$

for $\Gamma = \mathbf{B} + \mathbf{R}$.

Given estimates of $\Delta = cov(f) - cov(a)$ and $cov(o-b)$ from sample data, an estimate of system uncertainty can be derived as solution of the equation above:

$$cov(q) = cov(d)[cov(d)^{-1}\Delta]^{1/2}$$

but this is an entwined way to estimate $\mathbf{Q}$, unlike the lag-1 residual approach.

Note in passing that

- The filter analyses are BLUE, that is, no other estimate has smaller errors: lag-1 retrospective forecasts have larger errors than filter analyses by a measure proportional to system errors.

Note

- Sure, the solution here should be symmetrized.

So

- It would seem, that at best CH would get an indirect estimate of Q, but would it really get it?

The Fundamental Assumption in CH Methods

- The fundamental assumption for CH methods to work is that of independence among its chosen datasets.
- The 3CH choice of {o,b,a} is odd in the sense that not all of its datasets are independent.
- Although observation and background errors are uncorrelated, errors in the analysis are not uncorrelated to those, but …
- Luckily, it turns out that to get R and B all that is needed is for the errors in the analysis to be orthogonal to the innovations. And this is exactly what happens in an optimal system.
- 3CH fails to get A. It gets its negative instead!
- From the items above, it would seem a stretch to replace the third corner with retrospective forecasts, {o,b,f}, and expect to get anything useful, but …

Then again, what does 3CH get from {o,b,f}?

Have 3CH choose $\{\mathcal{O},\mathcal{B},\mathcal{F}\}$ for its $\{\mathcal{X},\mathcal{Y},\mathcal{Z}\}$ datasets:

$$\begin{aligned}
\hat{\mathbf{X}} &= \frac{1}{2}\{cov(\mathbf{o}-\mathbf{b})+cov(\mathbf{o}-\mathbf{f})-cov(\mathbf{b}-\mathbf{f})\}\\
\hat{\mathbf{Y}} &= \frac{1}{2}\{cov(\mathbf{b}-\mathbf{f})+cov(\mathbf{b}-\mathbf{o})-cov(\mathbf{f}-\mathbf{o})\}\\
\hat{\mathbf{Z}} &= \frac{1}{2}\{cov(\mathbf{f}-\mathbf{o})+cov(\mathbf{f}-\mathbf{b})-cov(\mathbf{o}-\mathbf{b})\}
\end{aligned}$$

Interestingly, it can be shown that the 3CH estimates for observation, background and retrospective forecast errors are:

$$\begin{aligned}
\tilde{\mathbf{X}} &\overset{opt}{=} \mathbf{R}+\mathbf{Q}\\
\tilde{\mathbf{Y}} &\overset{opt}{=} \mathbf{B}-\mathbf{Q}\\
\tilde{\mathbf{Z}} &\overset{opt}{=} \mathbf{Q}\Gamma^{-1}\mathbf{Q}-\Delta
\end{aligned}$$

where

$$\begin{aligned}
\Delta = \frac{1}{2}\{&\mathbf{R}\Gamma^{-1}\mathbf{Q}+\mathbf{Q}\Gamma^{-1}\mathbf{R}\\
&-\mathbf{R}\Gamma^{-1}\mathbf{B}-\mathbf{B}\Gamma^{-1}\mathbf{R}\\
&-\mathbf{B}\Gamma^{-1}\mathbf{Q}-\mathbf{Q}\Gamma^{-1}\mathbf{B}\}
\end{aligned}$$

When 3CH choses $\{\mathbf{o},\mathbf{b},\mathbf{f}\}$, none of the neglected terms in 3CH vanish. That is,

$$\begin{aligned}
\Delta\mathbf{X} &= E(\epsilon^o\odot\epsilon^b)+E[\epsilon^f\odot(\epsilon^o-\epsilon^b)]\\
\Delta\mathbf{Y} &= E(\epsilon^o\odot\epsilon^b)-E[\epsilon^f\odot(\epsilon^o-\epsilon^b)]\\
\Delta\mathbf{Z} &= E[\epsilon^f\odot(\epsilon^o+\epsilon^b)]-E(\epsilon^o\odot\epsilon^b)
\end{aligned}$$

- Uncorrelated O & B errors: $E(\epsilon^o\odot\epsilon^b)=\mathbf{0}$
- But $E[\epsilon^f\odot(\epsilon^o-\epsilon^b)]=-\mathbf{Q}\neq\mathbf{0}$

Therefore,

$$\begin{aligned}
\Delta\mathbf{X} &\overset{opt}{=} -\mathbf{Q}\\
\Delta\mathbf{Y} &\overset{opt}{=} \mathbf{Q}\\
\Delta\mathbf{Z} &\overset{opt}{=} \mathbf{A}+\Delta
\end{aligned}$$

In this case, 3CH fails to get any of the corners right.

Fail on paper, but viable in practice

- When establishing the relationship between 3CH and DBPC we discovered that, under optimality, both procedures obtain the same estimates of observation and background errors.
- Although in that context we find 3CH to fail to get DBPC analysis error estimate, the 3CH estimate can be made useful by simply recognize that it gets the negative of the analysis error as opposed to DBCP.
- Similarly, although none of 3CH estimates are correct when the corners corresponding to observation, background and retrospective forecast. An estimate of system uncertainty can still be recovered by apply 3CH to the two sets {o,b,a} and {o,b,f} separately. Use of the estimates for R or B from the first set can be used to infer Q from two of the estimates of the second set, without having to solve a quadratic equation to infer Q.
- This two-shot 3CH should be equivalent to applying 4CH to {o,b,a,f} but in this case, it is likely that we'd need to solve a quadratic equation to infer Q … to be done.

Closing Remarks

- Following a recently established relationship between 3CH and DBCP, this work tries to apply similar rationale to establish the relationship between CH methods and a residual-based approach to a lag-1 smoother procedure to estimate system uncertainty (Q).

- It is found that, unlike the residual-based lag-1 smoother approach, CH methods cannot get a direct estimates of Q.

- However, with the newly established relationship it is possible to get an estimate for Q by using a two-tiered 3CH approach:

  - Use {o,b,a} to get an estimate of R from the first corner, followed by
  - Use of {o,b,f} to getting an estimate of R + Q from the first corner, and thus derive Q.

- Obviously, the tricky part if getting retrospective forecasts f, but that is true of both approaches.