Response to Report 2 (reviewer 3):

General Reply: We thank the reviewer for the insightful remarks and hope that we could reply and adopt the manuscript in a sufficient way.

This study addresses the challenge of accurately determining error statistics for data assimilation and other uses. The work presented here builds on a growing body of geophysical applications of estimating error statistics using collocated observations. A detailed mathematical theory is presented to demonstrate how three or more data sets, with limited assumptions and no knowledge of truth, may be used to simultaneously and/or sequentially calculate error covariances and error cross-covariances for all data sets. The authors present illustrative examples of its application and a useful conceptual model for how it could be applied.

I find the work to be quite complex, in a way that I appreciate. That is, the level of detail is sufficient without being overbearing. And I believe this adds an intriguing method (or at least considerable detail to existing methods) to the literature on error estimation from multiple data sets. It's compelling to think on how one might apply the method to existing data, especially the determination of the three data sets that comprise the basic triangle and the subsequent order of the calculation. However, I think the authors have addressed this with sufficient detail for the present work.

I recommend a number of minor and technical edits to the text prior to publication.

Minor comments:

- **The use of scalar D_i for number of estimated error statistics and the matrix D_i;j for error dependency matrices and residual dependency matrices is unfortunate. I think it's clear enough, and perhaps this is common notation in the literature of which I'm not aware. But, the authors may consider changing the scalar D_i notation for clarity.**

  Reply: The scalar quantity D_I was never defined in literature, so there is no restriction to the use of a specific character. Actually, the choice of using of scalar D_I similar to the error dependency matrix $D_{i;j}$ was made on purpose because the scalar D_I it is the number of estimated error dependencies $D_{i;j}$ (of cross-covariances in general). But we agree that the double-use of D might be confusing, so we replaced the scalar D_I by A_I (refering to the number of **A**dditionally estimated error cross-statistics) though the manuscript, incl l.108, Eq.(3), Fig. 1, ...

- **Section 4.2 might be better titled as "Approximation for more than three datasets" as Section 4.1, which uses three datasets, already has "multiple" datasets.**

  Reply: Replaced as suggested (l.409).

- **Figs. 2, 3, and 4: the bottom axes of some of the subplots are cut off in this version. Please check that this is not the case in the final figures.**

  Reply: Thank you for the careful evaluation. However, there is no bottom axis at the third row of Fig. 2a,2b,3a,3b,4a,4b because the lower axis refers to the lower-right part of each subplot which in empty in the third row. During the 1st review, one reviewer got confused expecting the lower-right part of the third row to be zero-values. So we decided to remove the bottom axis in these subplots for clarification (among adding thin gray lines in the lower-right part). We now added a note on the empty lower-right subplots in the third row in caption of Fig.2 and in the description of the plots (ll.576-577):

" Note that the lower-right part of each subplot in the 3rd row does not contain any data. "

- **The "innocov" and "innocross" terms in algorithms A1 and A2 should be defined. They refer to the residual covariances and residual cross-covariances, correct? It's not clear with this naming, so please either change the terms or provide a (brief) definition. Are these perhaps terms from a previous version of the manuscript?**

  Reply: Thank you for this hint. Based on the 1st review, we renamed the innovation covariance and cross-covariance to residual covariances and cross-covariance in the manuscript, but forgot to update the algorithms. We replaced the terms accordingly in Alg. A1 and A2 and converted both algorithms to a template-conform format without changing its content.

- **Many instances of "were" being used in place of "where".**

  Reply: Thank you. It was corrected thought the whole document.

- **Line 15: remove comma between "both" and "background"**

  Reply: Done.

- **Line 24: "raises" instead of "arises"**

  Reply: Corrected.

- **Line 27: remove "being"**

  Reply: Done.

- **Line 30: remove "and" ("but has only")**

  Reply: Done.

- **Line 36: "optimality"?**

  Reply: Corrected.

- **Line 62: "where each approach"?**

  Reply: Corrected.

- **Line 69: spelling of "unknowns"**

  Reply: Corrected.

- **Line 114: suggest "as is the case"**

  Reply: Replaced with *"as in the case"* (l.114).

- **Fig. 1 caption: spelling of "and" (instead of "ans")**

  Reply: Corrected.

- **Line 257: suggest "other datasets and a"**

  Reply: Corrected.

- **Line 282: spelling of "datasets"**

  Reply: Corrected, also at all other locations in the manuscript. Thank you.

- **Line 318: should this be "D_i" instead of "U_i"?**

  Reply: Yes. Corrected, thank you.

- **Line 405, 448: perhaps "partial independence"?**

  Reply: Corrected.

- **Line 430: spelling of "cross-covariances"**

  Reply: Corrected.

- **Line 489: suggest "20,000" (comma instead of period)**

  Reply: Corrected.

- **Line 520: suggest "basic triangle are illustrated"**

  Reply: Corrected.

- **Fig. 3 caption: should be "but with a neglected"**

  Reply: Corrected.

- **Line 563: spelling of "solvability"**

  Reply: Corrected.

- **Fig. 6 caption: would it be better to state the distance as representing "the accuracy of assumed independence "?**

  Reply: Fig.5 and 6 are defined in the general case where error dependencies (or error cross-covariances) are assumed in some way (compare caption of Fig.5). In this context, the distance of dataset in Fig.6 indicates the accuracy of the related dependency; where the independence assumption is one choice of setting the assumed error dependencies to zero. Therefor, we kept the current formulation.

- **Line 588: should be "affect" rather than "effect"**

  Reply: Corrected.

- **Line 639: "major challenge"**

  Reply: Corrected.

Response to Report 3 (reviewer 2):

General Reply: We thank the reviewer for the very thoughtful and detailed evaluation and valuable remarks. We hope that we could reply and adopt the manuscript in a sufficient way.

Before providing detailed replies to each comment below, the main modifications of the manuscript are summarized here (compare replies to individual comments for more details):

- reformulation of Sec.5 for clarification of experiments (compare report3, comment2)
- reformulation of Sec.6.2, removing the claim for optimality of the solution (compare report3, comment1)
- reformulation and extension of Sec.6.1, placing the claim for requirement of the two conditions by a conceptual description of their sufficiency to solve the problem (compare report3, comment5)
- extension of the triangular estimate of error covariances to a polynomial estimate mainly in Sec.3.3.1, 4.2.1, 4.2.4, 4.2.5 (compare report3, comment5)

From our point of view these changes improved the manuscript significantly without changing its main statements.

This new version of the paper is for me a substantial improvement. I had written in my previous review that I was confused by the way the paper was written and that I had difficulties in understanding what the authors had exactly done. I think I have now understood (expect for Section 5, see comment 2 below, and for possibly minor details). That is due to improvements in the paper (for instance, in agreement with the first comment made by R. Todling in his review, the systematic use of the word *residual* instead of *innovation*). It is also due to the fact that a second reading of a paper, after some time, very often leads to a clearer understanding.

The paper is original and instructive and may be useful for many applications, for instance, as mentioned by the authors in their conclusion, for the combined use of analyses or forecasts produced by different Numerical Weather Prediction centres. On the other hand, I think improvement is still necessary, not in the scientific content of that paper, but in the way it is written and in the conclusions that the authors draw from their results. My main comments and suggestions, in approximate order of decreasing importance, are given below. I could have included some of these in my previous review, but I did not either because I had not understood some aspects I have now understood better, or because I considered these comments and suggestions of secondary importance at that stage. My main two comments (points 1 and 2 respectively) bear on the optimality that the authors claim to have defined and on the presentation of the numerical results (Section 5).

1. **The authors write (abstract, ll. 7-8) that their paper provides a *formulation of the minimal and optimal conditions to solve the problem* [i.e. the problem of estimating the statistics of the errors affecting collocated data] (see also section 6.2 *Optimal setup*, and statements on e.g. ll. 84, 544, 559, 645).**

   The minimal condition to solve the considered estimation problem is to define hypotheses that make that problem exactly determined. The authors indeed define an approach that satisfies that condition.

But concerning optimality, the above claim is largely exaggerated. I understand the authors mean that the uncertainty in the estimated statistics is in some sense minimum, but give no precise criterion by which criterion the corresponding optimality is defined. The 'optimal setup' is schematically described by Figure 6, and is based on a priori (and largely subjective) hypotheses as to the degree of correlation of the errors in the various datasets. That is by no means 'optimal' in any precise sense.

The sentence (ll. 600-601) *Thus, multiple independence trees can be defined* ... clearly says that different choices of triangles and reference subsets can be defined, which cannot be all 'optimal'.

In addition, the authors state clearly (l. 627) that positive definiteness of the estimated covariance matrices might not be fulfilled with their approach. Positive definiteness is obviously necessary for 'optimality' of estimates of variances and covariances.

I think the authors should either state precisely by which measure they consider their approach is optimal, or (preferably) remove any claim of optimality.

Reply: We agree with the reviewers concern about the use of the word "optimal" in the context of the setup of the problem. As noticed by the reviewer, this mainly refers to Sec.6.2 "optimal setup" and some other locations in the manuscript which refer to this section. Actually, this section aims to provide some guidelines for the selection of a setup among the possible setups which were determined in Sec.6.1 "minimal conditions". Given the large number of possible setups - which actually also include the definition of multiple independence trees in the reviewers reference to ll. 600 (old count) - we think that it is important to stress that there might be significant differences in real applications were assumptions remain inaccurate and that it is therefore crucial to select an appropriate setup. The guidelines in this section only refer to the aim of "minimizing" the uncertainties introduced by assumptions. The reviewer is right that positive definiteness is one another important criterion which would be required for truly optimal covariance estimates.

We extended the introduction of Sec.6.2 for clarification (ll.724-729):
" The general rules given in Sect. 6.1 allow for multiple different setups of datsets which all solve the error estimation problem. However in real applications, theremight be significant differences in estimated error statistics fromdifferent setups as observed e.g. by Vogelzang and Stoffelen (2021) in the scalar case. The optimal selection is specific for each application and may depend on several requirements related to the actual purpose or use (e.g. available knowledge, need for positive definiteness, accuracy of each estimate). This section provides some general guidelines on the selection of an appropriate setup among the various possible solutions w.r.t. the uncertainties introduced by statistical assumptions. "

We also went though the manuscript and reformulated all locations where the word "optimal" appeared in the context of the setup of the problem:

- abstract (l.7): *"guidelines for setting up and solving the problem"*
- introduction (ll.65-66): *"And what are the general conditions to set up and solve the problem?"*
- and (ll.83) *"and provides guidelines for the setup of those"*
- title of Sec.6 (l.667): *"Conceptual summary and guidelines"*
- introduction of Sec.6 (l;.670-671): *"... Sect. 6.2 formulates guidelines for the selection of an appropriate setup of datasets under imperfect assumptions."*
- Sec.6.1 (l.557-561, old count) removed
- title of Sec.6.2 (l.723): *"Selection of setup"*
- Sec.6.2 (ll.735-736): *"In order to achieve sufficiently accurate estimates, ... "*

- and (l.743): *"..., the setup visualized in Fig. 5 is not an appropriate selection."*
- conclusions (l.645): *"optimal"* removed.

2. **I find Section 5, which presents results of numerical experiments, rather confusing and difficult to understand. The experiments that have been performed are not really described. What were for each experiment the real error statistics that were used for producing the datasets and then computing the residual statistics ? What where then the error statistics that were a priori assumed in order to close the problem of estimation of the global error statistics ? Were those a priori assumed statistics consistent with the real ones ? I understand that was the case for the experiment whose results are presented in Fig. 2a, but not for the other ones, although that is not clearly said. If they were consistent, were all the a posteriori estimated statistics in agreement with the real ones ? I understand that was the case for Fig. 2a, (... *the two remaining dependencies are estimated accurately*, l. 511), with the consequence that the 'matrices' in the two bottom rows of Fig. 2a must be exactly symmetric, with in particular the matrices in the third row being full of zeroes. That seems visually to be the case, but is not mentioned explicitly, leaving the reader in some doubt.**

   In the other experiments, the assumed statistics were not consistent with the real ones. That was the case for Fig. 2b, about which the authors write (ll. 525-526) *As shown in Fig. 2b, the error covariance of dataset 4 is underestimated by half the neglected dependency between (2;4)*. How can that (including the quantitative assessment) be seen from Fig. 2b ? And how can the reader check that the errors in the a posteriori estimated statistics are themselves consistent with the corresponding estimations presented in Section 4 ? The reference to Eq. (42), (51) and (52) (l. 521) is not of much help.

   Considering figures, they must help then reader, who must be able to distinguish, among the statements made by the authors, what can be seen in the figures, and what cannot (in the latter case, the simple mention *not shown* is necessary).

   These are only examples. I do not suspect that anything is scientifically wrong or disputable in Section 5 (nor anywhere else in the paper for that matter). But I think the Section must rewritten, with a precise description of each of the experiments that have been performed, with explicit statement of the associated hypotheses, and with a precise description of the obtained results, as well as of the conclusions that must be drawn from these results.

   The only difficulty left to the readers must be with a proper understanding of the approach taken by the authors and of what the latter have done. Anything that has to do with deciphering the figures and drawing the conclusions must require no additional effort from the reader.

   Reply: Thank you for this detailed comment, we understand that the Section was not clear enough. We reformulated Sec. 5, focusing especially on a clear introduction and description of the experiments, including the formulation of the setup, the related assumptions and the estimation procedure. The description of the results in Subsec.5.1 to 5.3 was also reformulated to assist the reader in the interpretation. We tried to make clear where the descried results can actually be seen in the figures (the references to specific panels are given in brackets) and also give some references to the referring equations in the theoretical sections Sec.3 and 4. This also improves the connection of this section to the rest of the manuscript. We also extended the formulation of conclusions of the experiment. We believe that this improves the readability of the Section significantly and hope that it answers all of the reviewers questions sufficiently.

Because this reformulation increased the length of Sec.5, we decided to have a separate subsection for each of the three experiments, i.e. divide Sec.5.2 into two subsections.

The modification affects the complete Sec.5, which we did not copy here. Instead, the general changes are listed below:

- reformulation introduction of Sec.5 for clarification (ll.534-567)
- addition of paragraph in each subsection which introduces the experiments (ll.579-584, ll.605-610, ll.639-642)
- reformulation results of each subsection (ll.585-596, ll.611-634, ll.642-661)
- division of Sec.5.2 into two subsections; rename Sec.5.2 for experiment 2 *"Small uncertainties in basic triangle"* (l.604) and define new Sub.5.3 for experiment 3 *"Large uncertainties in basic triangle"* (l.633)
- addition of a concluding paragraph in each subsection (ll.597-603, ll.635-637, ll.661-666)
- addition *"Experiment XY:"* in captions of Fig.2,3,4
- extension of figure captions (i.e. caption Fig.2): *" For each subplot, gray asterisks in the upper-left part of the first row indicate that these error dependencies are assumed to be zero in the estimation. Note that the lower-right part of each subplot in the third row does not contain any data. "*

3. **As already mentioned in my previous review, Eq. (23) can be obtained directly without going through the error statistics C, X or D. The authors write in their response that they have rearranged the equation to show the equivalence between innovation and error statistics in a clearer way. My point is that there is simply no need to refer to a 'truth' or to 'errors' in order to obtain Eq. (23). The equivalence between innovation and error statistics has been clearly shown by Eqs (19-22). Introducing quantities in places where there are useless can only confuse the reader (the fact there is a link between innovation and error statistics is irrelevant at this precise point).**

   Reply: Thank you for the clarification. We now understand the reviewers point and show the equivalence in Eq. (23) directly from the definitions - which is actually more consistent to the rest of Sec.3.2 - without using the true state or error statistics (ll.222-228 and Eq.(23)):
   " For $k = i$, the residual dependency between $i - j$ and $i - l$ can be expressed as combination of three residual covariances: "

$$
\begin{aligned}
\mathbf{\Gamma}_{i;l} + \mathbf{\Gamma}_{j;i} - \mathbf{\Gamma}_{j;l} &\overset{(6)}{=} \overline{\left[x_i(p) - x_l(p)\right]\left[x_i(q) - x_l(q)\right]} + \overline{\left[x_j(p) - x_i(p)\right]\left[x_j(q) - x_i(q)\right]} - \overline{\left[x_j(p) - x_l(p)\right]\left[x_j(q)\right.}\\
&= \overline{\left[x_i(p)\right]\left[x_i(q)\right]} - \overline{\left[x_i(p)\right]\left[x_l(q)\right]} - \overline{\left[x_l(p)\right]\left[x_i(q)\right]} + \overline{\left[x_l(p)\right]\left[x_l(q)\right]}\\
&\quad + \overline{\left[x_j(p)\right]\left[x_j(q)\right]} - \overline{\left[x_j(p)\right]\left[x_i(q)\right]} - \overline{\left[x_i(p)\right]\left[x_j(q)\right]} + \overline{\left[x_i(p)\right]\left[x_i(q)\right]}\\
&\quad - \overline{\left[x_j(p)\right]\left[x_j(q)\right]} + \overline{\left[x_j(p)\right]\left[x_l(q)\right]} + \overline{\left[x_l(p)\right]\left[x_j(q)\right]} - \overline{\left[x_l(p)\right]\left[x_l(q)\right]}\\
&= \overline{\left[x_i(p) - x_j(p)\right]\left[x_i(q) - x_l(q)\right]} + \overline{\left[x_i(p) - x_l(p)\right]\left[x_i(q) - x_j(q)\right]}\\
&\overset{(4)}{=} \mathbf{\Gamma}_{i-j;i-l} + \mathbf{\Gamma}_{i-l;i-j}
\end{aligned}
$$

4. **Ll. 303-304, *The equivalence demonstrates that the exact formulations of error statistics from residual covariances and cross-covariances are consistent to each other* (incidentally, the correct wording would be *consistent with each other*). I understand the authors stress the 'equivalence' of those formulations because they will show later that,**

under the assumption of 'independence', they may lead to different results. But the reader cannot understand at this stage why it is useful to stress the equivalence. I suggest the authors write rather *The equivalence demonstrates that, as they must be, the exact formulations* .... And the equivalence does not ultimately result from the fact that Eq. (20) is a special case of Eq. (22) (ll. 305-306), but from the basic definitions (4) and (5). A similar argument holds before that for Eq. (32) and the comment that follows. I suggest to write (l. 289) ... *the formulation of error covariances based on residual crosscovariances in Eq. (31) is, as it must be, symmetric and equivalent* ...

Reply: Corrected as suggested, with updated equation numbers (31)→(34) (ll.301-302): *" the formulation of error covariances based on residual cross-covariances in Eq. (34) is, as it must be, symmetric and equivalent to the formulation based on residual covariances from Eq. (25). "*
and (ll.315-318): *" The equivalence demonstrates that, as they must be, the exact formulations of error statistics from residual covariances and cross-covariances are consistent with each other. This consistency applies to the exact formulations of all symmetric error statistics (error covariances and dependencies) and results from the consistent definitions of residual covariances and cross-covariances in Eq. (4) and (6). "*

5. **Ll. 553-554, ... *there are two requirements for the setup of datasets* It does not seem to me that the need for either one of two stated requirements has been shown. It has only been shown that those two requirements are sufficient for solving the underlying estimation problem (actually, I understand the sentence ll. 569-570 as meaning that other possibilities exist).**

Reply: We understand that this point is not sufficiently clear in the manuscript. The two requirements are logical conclusions from the theoretical sections before (Sec. 3+4), but they do not provide a sufficient mathematical foundation that shows that there are not other possible solutions of the problem.

Indeed, it would be very interesting to approach the closure problem (i.e. solvability, optimality, ...) from a purely mathematically point of view. However, we believe that a rigorous mathematical proof exceeds the scope of this journal and the purpose (and length) of this paper. We hope to address this problem from a purely mathematical point-of-view in an upcoming study.

A) Here, we decided to give a logical argumentation why these two "requirements" are sufficient solutions of the problem without claiming the necessary need of those. The specific modifications to the manuscript are as follows:

- We replaced the word "requirement" by "rules" and the claim for the need for the two "requirements" by a statement that these "rules" are sufficient for solving the problem (l.669): *"Section 6.1 summarises the general assumptions and provides rules for the minimal conditions ... "*
- Sec.6.1 was reformulated to guide the reader though the argumentation which leads to the two "requirements" (now "rules"). Note that these changes already include the modifications referring to part B of this reply.
  - (ll.674-676): *" This section provides a conceptual discussion of different conditions which need to be fulfilled in order to be able to solve the error estimation problem. The discussion is based on the previous sections, but formulated in a qualitative way without providing mathematical details. "*
  - (ll.684-686): *" The number of error statistics that can be estimated for a given number of datasets (N_I) was introduced in Sect. 2. However not every possible combination of error statistics to be estimated provides a solution. The following discussion only considers setups where all error covariances and as many error cross-statistics as possible are estimated. "*
  - (ll.684-705), logical argumentation (see manuscript)
  - (ll.706-708): *" Based on this, two general rules for the setup of datasets can be formulated which ensure the solvability of the problem in the case where all error covariances and as many error cross-statisitcs as possible (cross-covariances or dependencies) are estimated: ... "*
  - remove statements on requirements which were below l.713 (new count)

- other locations in the manuscript which refer to the claim for minimal requirements were also modified accordingly:
  - Sec.1 Introduction (ll.82-83): *" It includes the formulation and illustration of rules to solve the problem for an arbitrary number of datasets and provides guidelines for the setup of those. "*
  - Sec.2 General framework (l.105): *" Sect. 6.1 provides guidelines which ensure the solvability of the problem for a minimal number of assumptions. "*

For clarification, the "illustrative example" (reformulated in accordance to comment below) is an example setup to demonstrate which these two requirements could be fulfilled. It is one option of the large amount of different possible setups which fulfill the requirements and is not meant to indicate that other possibilities exist which would not fulfill the requirements. We added a note to clarify this (ll.718-719):
" Note that this is one of many possible setups which are determined by the two rules above. "

B) Related to these changes, we noticed that the current formulation misses one important generalization, which actually defines a larger group of possible solutions of the estimation problem. Up to now, the "direct estimation" of error covariances, with no other error covariances available, was only created by a "triangular estimation" from the combination of 3 residual covariances (eg. Eq. (27)). However, this can be generalized to a "polynomial estimation" by combining a closed series of F residual covariances, for any F uneven and larger or equal 3 (and smaller or equal the number of datasets I – see the new parts in the manuscript for a detailed description). We believe that this generalization is an important aspect which needs to be considered in the manuscript and adopted the manuscript accordingly.

This generalization does not change the main content of the study but required a number of minor modifications thought the manuscript, most of them related to the replacement of words, e.g. replacing "triangular estimate" and "basic triangle" by "polynomial estimate" and "basic polynom", respectively. The most significant modifications are the addition of the generalized equations for:

- the generalized exact formulation of error covariances from residual covariances in Sec.3.3.1 (ll.269-277): *" Equation (26) can be generalized by replacing the closed series of the 3 dataset-pairs $(i;j),(j;k),(k;i)$ with a closed series of F dataset-pairs $(i_1;i_2),(i_2;i_3),\cdots,(i_{F-1};i_F),(i_F;i_1)$, for any $3 \leq F \leq I$ (and I the number of datasets):*

$$\mathbf{C}_{\tilde{i_1}} \overset{(20)}{=} \sum_{f=1}^{F-1}(-1)^{f-1}\left[\mathbf{\Gamma}_{i_f;i_{f+1}}+\mathbf{D}_{\tilde{i_f};\tilde{i_{f+1}}}\right]+(-1)^{F-1}\left[\mathbf{\Gamma}_{i_F;i_1}+\mathbf{D}_{\tilde{i_F};\tilde{i_1}}\right]+(-1)^{F}\mathbf{C}_{\tilde{i_1}}$$

  Because of changing signs, Eq. (28) can only be solved for the error covariance $\mathbf{C}_{\tilde{i_1}}$ if F is uneven. If F is even, $\mathbf{C}_{\tilde{i_1}}$ cancels out and cannot be eliminated, and Eq. (28) could be solved for one error dependency instead. If F is uneven, the generalized formulation for $\mathbf{C}_{\tilde{i_1}}$ becomes:

$$\mathbf{C}_{\tilde{i_1}} \overset{(28)}{=} \frac{1}{2}\left[\underbrace{\left(\sum_{f=1}^{F-1}(-1)^{f-1}\mathbf{\Gamma}_{i_f;i_{f+1}}\right)+\mathbf{\Gamma}_{i_F;i_1}}_{\textit{"independent contribution"}}+\underbrace{\left(\sum_{f=1}^{F-1}(-1)^{f-1}\mathbf{D}_{\tilde{i_f};\tilde{i_{f+1}}}\right)+\mathbf{D}_{\tilde{i_F};\tilde{i_1}}}_{\textit{"dependent contribution"}}\right] \quad , \forall\ F \text{ uneven} \wedge 3 \leq F$$

  where Eq. (27) results for $F = 3$ with indices $i_1 = i$, $i_2 = j$ and $i_3 = k$. Note that in any case, the number of assumed and estimated error statistics remains consistent with the general framework in Sect. 2. "

- the generalized approximative form for direct estimation of error covariances from residual covariances (new Sec.4.2.1, ll.424-432):
  " For more than three datasets $I > 3$, the estimation from three residual covariances in Eq. (39) can be generalized to estimations of error covariances from a closed series of F residual covariances (compare Sect. 3.3.1). For any uneven F with $3 \leq F \leq I$, each error covariance can be estimated under the assumption of vanishing error dependencies along the closed series of datasets $\mathbf{D}_{\tilde{i_f};\tilde{i_{f+1}}}\ \forall\ f \in [1, F-1]$

and $\mathbf{D}_{\tilde{i_F};\tilde{1}}$*:*

$$\mathbf{C}_{\tilde{i_1}} \underset{\{inF\}}{\overset{(28)}{\approx}} \frac{1}{2}\left[\left(\sum_{f=1}^{F-1}(-1)^{f-1}\mathbf{\Gamma}_{i_f;i_{f+1}}\right)+\mathbf{\Gamma}_{i_F;i_1}\right] \qquad , \ \forall\ F\ uneven\ \wedge\ 3\leq F\leq I$$

where " $\underset{\{inF\}}{\approx}$ *" indicates the assumption of neglectable error dependencies along the series of datasets. As shown in Sect. 2, the problem cannot be closed for less than 3 datasets, even under the independent assumption. For* $F=3$ *datasets, Eq. (39) is a special case of Eq. (48) with indices* $i_1=i$, $i_2=j$ *and* $i_3=k$. *"*

- the uncertainty of the generalized direct error covariance estimate in Sec.4.2.4 (ll.464-468):
  " As generalization of Eq. (45), the absolute uncertainty $\Delta\mathbf{C}_{\tilde{i_1}}$ *of a polynomial error covariance estimate introduced by the assumption of pairwise-independence along the closed series of* F *datasets, with* F *uneven and* $3\leq F\leq I$*, is given by:*

$$\Delta\mathbf{C}_{\tilde{i_1}}\Big|_{(47)} := \mathbf{C}_{\tilde{i_1}}\Big|_{true} - \mathbf{C}_{\tilde{i_1}}\Big|_{(47)} \overset{(28),(47)}{=} \frac{1}{2}\left[\left(\sum_{f=1}^{F-1}(-1)^{f-1}\Delta\mathbf{D}_{\tilde{i_f};\tilde{i_{f+1}}}\right)+\Delta\mathbf{D}_{\tilde{i_F};\tilde{i_1}}\right] \qquad , \ \forall\ F\ uneven\ \wedge\ 3\leq$$

Due to the changing sign of error dependencies along the series of datasets, the absolute uncertainty of the error covariance estimates does not necessary increase with the size of the polygon F. *"*

- the equivalence of the generalized direct error covariance estimate in Sec.4.2.5 (ll.505-512):
  " This can also be generalized for the estimation of any error covariance $\mathbf{C}_{\tilde{i_2}}\big|_{\vdash}$ *given its reference* $\mathbf{C}_{\tilde{i_1}}\big|_{\pentagon}$ *estimated with the polygonal formulation for a closed series of* F *pairwise-independent datasets for any uneven* F *with* $3\leq F\leq I$*:*

$$\begin{aligned}
\mathbf{C}_{\tilde{i_2}}\big|_{\vdash} &\underset{\{inI\}}{\overset{(48)}{\approx}} \mathbf{\Gamma}_{i_1;i_2} - \mathbf{C}_{\tilde{i_1}}\big|_{\pentagon} \\
&\underset{\{inF\}}{\overset{(47)_{i_1}}{\approx}} \mathbf{\Gamma}_{i_1;i_2} - \frac{1}{2}\left[\left(\sum_{f=1}^{F-1}(-1)^{f-1}\mathbf{\Gamma}_{i_f;i_{f+1}}\right)+\mathbf{\Gamma}_{i_F;i_1}\right] \\
&= \frac{1}{2}\left[\left(\sum_{f=2}^{F-1}(-1)^{f-2}\mathbf{\Gamma}_{i_f;i_{f+1}}\right)-\mathbf{\Gamma}_{i_F;i_1}+\mathbf{\Gamma}_{i_1;i_2}\right] \quad \underset{\{inF\}}{\overset{(47)_{i_2}}{\approx}} \mathbf{C}_{\tilde{i_2}}\big|_{\pentagon} \qquad , \ \forall\ F\ uneven\ \wedge\ 3\leq F\leq I
\end{aligned}$$

The consistency between direct and sequential error covariance estimates results directly from their common underlying definition of residual covariances in Eq. (20) and holds not only for the approximate formulations but similarly for the full expressions including error dependencies (compare Sect. 3.3.1). "

The remaining modification are as follows:

- reformulation in abstract (ll.10-12): *" The presented generalized estimation of full error covariance- and cross-covariance matrices between dataset does not necessarily accumulate uncertainties of assumptions among error estimations of multiple datasets. "*
- reformulation of the general introduction of Sec.4 (ll.322-324): *" An extension for more than three datasets based on a minimal number of assumptions is introduced in Sect. 4.2. It includes the estimation of additional error covariances, either directly or sequentially, and some error cross-statistics to estimate a maximum amount of error statistics. "*
  and (ll.327-328): *" ... and (iii) the comparison of the approximations from direct and sequential estimates (Sect. 4.2.5). "*
- reformulate introduction of Sec.4.2 (ll.401-407): *" As described in Sect. 2,* $A_I>0$ *gives the number of error cross-statistics which can potentially be estimated in addition to all error covariances for* $I>3$ *datasets. Consequentially, the independent assumption between all pairs of datasets can be relaxed to*

a "partial independence assumption" where one independent dataset-pair is required for each dataset I. The estimation of error covariances can be generalized in two ways: Firstly, the direct formulation for three datasets in Sect. 4.1.1 is generalized to a direct estimation of more than three datasets in Sect. 4.2.1. Secondly, Sect. 4.2.2 introduces the sequential estimation of error covariances of any additional dataset. This estimation procedure of additional error covariances is denoted as "sequential estimation" because is requires the error covariance estimate of a prior dataset, in contrast to the "direct estimation" from an independent triplet of datasets ("triangular estimation" in Sect. 4.1) or generally from a closed series of pairwise-independent datasets ("polygonal estimation" in Sect. 4.2.1). "

- generalized formulation of the use of a basis pentagon as basis for the sequential covariance estimate in Sec.4.2.2 (ll.436-438): *" Similarly, a "basic polygon" can be defined from a closed series of F pairwise-independent datasets, where the referring error covariances can be directly estimated from Eq. (48). "*
- moving aspects of the sequential estimation from the general introduction of Sec.4.2 to Sec.4.2.2 (l.439-442): *" For each additional dataset i with $F < i < I$, its cross-statistics to one prior dataset ref(i) $< i$ is needed to be assumed in order to close the problem. This prior dataset ref(i) is denoted as "reference dataset" of dataset i. With this, the remaining error covariances can be estimated from residual covariances under the partial independence assumption $\mathbf{X}_{\tilde{i};\tilde{ref}(i)} = 0$. "*
- use index "G" instead of "F" for the sequential uncertainty of error covaiances to avoid confusion (ll.477-480)
- note on generalization of comparison to sequential estimate in Sec.4.2.5 (ll.527-529): *This holds similarly for any polygonal estimation, where the additional independence assumption which closes the series of pairwise-independent datasets has to be of similar accuracy as the other independent assumptions. "*
- generalization formulation of the 1st rule for the setup of datasets in Sec. 6.1 (ll.709-711): *" all error cross-statistics along a closed series of dataset-pairs, with the number of involved datasets uneven and larger or equal three, are needed (this closed series of datasets is called "basic polygon" or "basic triangle" in case of three datasets), and ... "*
- addition of a basic pentagon in the illustrative example: Fig.5(plot + title + title of Fig.6 accordingly) + (ll.721-722): *" Alternatively, the basic triangle could be replaced e.g. by a basic polygon of five datasets ("basic pentagon": 1;2;3;5;4), if the dependency(3;5) is assumed instead of the dependency(3;1). "*
- generalization of the general rules for the selection of a setup: Sec.6.2 (ll.733-735): *" Because uncertainties in error estimate do not necessarily sum up for a large basic polynom or along a branch of the independence tree (compare Sect. 4.2.4), a large residual-to-dependency ratio w.r.t. of the assumed cross-statistics is more important than a low number of intermediate datasets. "*
- generalization of Algorithm for residual covariances in Apx.A: Alg.A1(title and algorithm) + title of Alg.A2 + (ll.817-818): *" In this algorithm, the generalized formulation of a basic polygon of $F \leq I$ residuals, for any uneven $F \geq 3$, is used for the estimation of the first error covariance. "*
  + (ll.821-822): *" This algorithm uses a basic triangle as example for a basic polygon for the estimation first error covariance. "* + (ll.831-832): *" Note that the generalized basic polygon can also be used for the estimation of the first error covariance in Algorithm 2. "*
- (minor) addition of the word "sequential" when refering to sequential error covariance estimates in contrast to the generalized direct estimates: Sec.4.2.2 (l.433) + (l.439) + Sec.4.2.4 (l.469) + (l.473) + (l.483)
- (minor) replacement of "3" direct estimates by "F": Sec.4.2.4 (l.469) + Sec.4.2.5 (l.516)
- (minor) replacement of "basic triangle" to "basic polygon" and "independent triangle" by "pariwise-independent polygon": Sec.4.2.4 (l.476) + (ll.478-479) + Sec.4.2.5 (ll.514-515) + (l.518) + Sec.6.2 (ll.745-746) + (l.748) + (l.750) + (ll.752-754) + (l.758) + (l.760) + (l.752) + Apx.A (l.810)

6. **L. 552, ... *without introducing additional degrees of freedom.* That formulation is confusing. From what I understand, the purpose is fundamentally to eliminate degrees of freedom by**

introducing hypotheses that render exactly determined the problem of estimating the error statistics. It seems that you implicitly anticipate on the text (ll. 602-608) where it is suggested to combine different estimates.

Reply: This information was added to emphasize the following: If an assumed error statistic is formulated as function of an additional unknown, this additional unknown cannot be determined in the estimation scheme as it induces an additional degree-of-freedom which makes the problem underdetermined. We see that this formulation might actually be confusing here, we removed it form the sentence (ll.681-682):
" The only restriction is that all assumed error statistics must be fully determined by other error statistics or predefined values. "

7. **Ll. 551-552, *The only restriction is that all assumed error statistics must be fully determined* ... From what I understand, the wording all additional assumed error statistics would be more appropriate.**

   Reply: In this manuscript the word "additional" is referring to all datasets which are not part of the basic triangle (or basic poylgon). In this particular discussion it would also be possible to formulate e.g. an assumed error cross-covariance between two datasets in the basic triangle as function of their error covariances. Thus this sentence refers to all assumed error statistics, and not only the "additional" ones. Therefor, we decided to keep the current formulation.

8. **Concerning positive definiteness of the estimated error statistics, the authors write (ll. 393-394) *However, the generalization to covariances matrices is expected to increase the occurrence of negative values were* (incidentally, the proper spelling is *where*) *correlations between two entries of the state are low, thus relative differences and sampling errors become large.* I am not sure I understand what that means. Is it that the occurrence of non definite positiveness is likely to increase with the number I datasets, or what ?**

   Reply: This refers to the fact that - in comparison to scalar variances at each location - covariance matrices often contain small values where the spatial correlation between eg two spatially distant locations is small. This results in a small covariance between the two locations, and uncertainties in the assumptions (eg independence) or sampling noise can become larger than the actual true covariance which may lead to negative values in the estimated covariance at these locations. We reformulated the sentence for clarification (ll.406-408):
   " However, the generalization to covariances matrices is expected to increase the occurrence of negative values in off-diagonal elements. Because spatial correlations and thus true covariances may become small compared to uncertainties in the assumptions or sampling noise, estimated covariances at these locations might become negative. "

9. **L. 370, *All three estimates become equivalent if the residual cross-covariances are symmetric* Yes, but it might be useful to mention here that these equivalent estimates may not be positive definite.**
   And l. 388, *Estimated error covariances might even contain negative values* ... You must mean *might even not be positive definite* ... .

   Reply: We added a refering note (ll.384-385):
   " However, non of the estimates ensures positive definiteness of the error covariances. "

   And reformulated as suggested (l.401):
   " Estimated error covariances might even not be positive definite if ... "

10. **Ll. 626-627, *While the presented method ensures symmetry of error covariances*, Not necessarily (Eqs 37 or 38 do not ensure symmetry). Say that symmetry can be enforced if necessary.**

    Reply: Right. We reformulated for clarification (l.785):
    " While the presented method can be formulated to provide symmetric error covariances, ... "

Additionally, we also added this point in the description of the Algorithms in Apx.A (ll.827-831):
" The decision to estimate error statistics from residual covariances (Algorithm A1) or cross-covariances (Algorithm A2) depends on the availability of residual statistics, the need for symmetric estimations of error covariances – which is only intrinsically guaranteed in Algorithm A1 –, and the need for estimating asymmetric error cross-covariances – which can only be estimated with Algorithm A2 (compare Sect. 3.3.1). "

11. **L. 2 (abstract), …** ***an ill-posed problem*** **… Not** ***ill-posed*****, but** ***underdetermined***

    Reply: Replaced as suggested.

12. **Ll. 325-326,** ***"assumption of independence"*****. I mention that the word independence is not used here in its accepted standard meaning in probability theory. The hypothesis made here is actually an hypothesis of no correlation, which is a weaker property than independence. I suggest the authors briefly mention that fact.**
    And l. 325, ***equals*** **is to be changed to** ***implies***

    Reply: Thank you for this important point. We reformulated the sentence for clarification and replaced the word as suggested (ll.337-338):
    " Because the assumption of zero cross-covariance implies zero error correlation which is often used as proxy for independence, it is denoted as "assumption of independence" or "independence assumption" thereafter. "

13. **Ll. 440-441, …** ***absolute uncertainties of estimations from residual covariances and cross-covariances differ only in the uncertainties w.r.t. the basic triangle*** **… Well, these uncertainties depend also in the uncertainty in the hypothesis $\mathbf{D}_{i;ref(i)} = 0$.**

    Reply: That's right, the mentioned uncertainties include also $\mathbf{D}_{i;ref(i)} = 0$. However, the uncertainty induced by this assumption is the same among the two estimations (Eq. (54),(55)). The sentence refers only to the differences of the two estimations: "absolute uncertainties … differ only in … ".

14. **L. 315,** ***the determination of uncertainties resulting from possible errors in the ~~caused by~~ assumed error statistics***

    Reply: Corrected (l.327).

15. **L. 201, what are unbiased error statistics ? Error statistics are here the unknowns, not the data, and whether they are biased or not makes a priori no sense.**

    Reply: Corrected to "unbiased datasets" (l.201).

16. **L. 323, $\mathbf{X}_{i;j} = 0 \Leftrightarrow \mathbf{D}_{i;j} = 0$ to be changed to $\mathbf{X}_{i;j} = 0 \Rightarrow \mathbf{D}_{i;j} = 0$**
    L. 414, similarly, the condition $\mathbf{X}_{i;ref(j)} = 0$ is actually stronger than the assumption $\mathbf{D}_{i;ref(i)} = 0$ made on l. 405. Be more consistent as to which hypotheses you make.

    Reply: Thank you for pointing this out. We corrected l.335 (new count) as proposed and replaced $\mathbf{D}_{i;ref(i)}$ by $\mathbf{X}_{i;ref(i)}$ in l.442 (new count) as well as l.520 (new count). In addition, the description of the "independence assumption" was adopted accordingly, now referring to vanishing error cross-covariance $\mathbf{X}_{i;ref(i)}$ (compare comment 12 above). With this correction, the subsequent formulations e.g. in l.343 (new count) and l.444 (new count), remain correct. We checked the consistency thought the whole manuscript. The "independence assumption" now refers solely to vanishing error cross-covariance $\mathbf{X}_{i;j}$ which induces vanishing error dependencies $\mathbf{D}_{i;j}$.

17. **L. 558, $I \leq 3$, you must mean $\mathbf{I > 3}$?**

    Reply: Replaced by $I \geq 3$. Thank you.

18. **L. 318, $U_I \geq 0 \to D_I \geq 0$**

    Reply: Corrected.

19. **L. 352,** ... ***four equivalent formulations for each pair of other datasets*** **... . You have assumed here I = 3, so that there is only one pair of 'other datasets' (see also l. 358).**

    Reply: Right, this is an artefact from a more general description. We removed "for each pair of other datasets" in both cases (l.363, l.370):
    " ... every error covariance from residual cross-covariances has four equivalent formulations which provide the same result in the exact case ... "
    " Equations (36) to (38) provide three different estimates of an error covariance matrix. "

20. **L. 34, what are exactly the *corners* ?**

    Reply: The "corners" refer to the three datasets when used in the (G)3CH method. We see that this remains unclear for readers who are not fimiliar with this method. We reformulated avoiding the used of the word "corners" (ll.34-35):
    " They show that when the G3CH method is applied to the observations, background and analysis of variational assimilation procedures, ... "

21. **Ll. 40-41, sentence starting *Up to now,* ... awkward. From what I understand, I suggest *Up to now, only scalar error variance estimation has been implemented in data assimilation with the TC method (e.g.* ...**

    Reply: Thank you, we reformulated based on the reviewers suggestion (ll.40-41):
    " Up to now, there are only a few applications of scalar error variance estimation in data assimilation with the TC method ... "

22. **The word *exemplary* is used mistakenly in several places (the word designates something that is meant to be imitated, while the authors obviously think of illustrative examples)**
    L. 543, ... ***including an exemplary visualisation,*** → ... ***including an illustrative visualisation,***
    L. 566, *An exemplary setup* ... → ***An illustrative example of setup*** ...
    Caption of Figure 5, *An illustrative example of visualization* ...
    See also l. 7 (abstract)

    Reply: Corrected at all 4 locations.

23. **Ll. 35-36,** ... ***this particular error estimation problem can only be closed under the assumption of optimally.*** **I do not understand what this means (and the proper wording should in any case be assumption of optimality)**

    Reply: The optimality (not optimally) refers to the analysis being optimal, which is a common assumption in a posteriori evaluation methods in data assimilation. We reformulated for clarification (ll.35-36):
    " ... this particular error estimation problem can only be closed under the assumption that the analysis is optimal. "

24. **Ll. 54-55 (and later) *error cross-variances* ... You must mean *error covariances* ?**

    Reply: According to the notation and definition of this manuscript, the error interaction between two scalar datasets is denoted as "cross-variance" because "cross-" always refers to the interaction between datasets (like residual or error cross-covariance, ll.150-156) whereas "variances" are the diagonal elements of "covariances". In the scalar case, which was e.g. formulated by Zwieback et al 2012, all covariances reduce to scalar variances. Thus, the term "cross-variances" is used when refering to scalar error interactions in order to be consistent with the notation in the rest of this manuscript. We extended the sentence to clarify this difference in wording (ll.54-55) which should also avoid misunderstanding lateron:
    " Zwieback et al.,2012 were the first proposing the additional estimation of the scalar error cross-variances between two selected datasets (which they denote as covariances) ...

25. **Ll. 65-66,** ... ***affect different formulations of the estimated error statistics?*** **→** ... ***affect different estimations of the error statistics?***

    Reply: Corrected.

26. **L. 613, I understand the N-CN *method* is what was called previously the *N-cornered hat method* (l. 603), or what ?**

    Reply: We assume that the reviewer refers to the use of "N-CH method" which is indeed the short form of the "N-cornered hat method". We added the abbreviation where the method was first mentioned (l.760) and the full name where the abbreviation is first mentioned in the conclusions (l.770).

27. **L. 24, *arises* → *raises***

    Reply: Corrected.

28. **L. 588, *effect* → *affect* (or *impact*)**

    Reply: Corrected to "affect".

29. **L. 589, *triple* → *triplet* (the same correction is to be made elsewhere, e.g. l. 407 ; please check)**

    Reply: Corrected everywhere, where it appeared.

30. **Ll. 574-575,** ... ***can be interpreted as being similar to*** ...

    Reply: Corrected.

31. **L. 583,** ... ***is comparably well known,*** ... **word *comparably* inappropriate here. I suggest** ... ***is known to some degree of accuracy,*** ......

    Reply: Corrected.

32. **L. 544,** ... ***algorithmic summary for the calculation*** ...

    Reply: Corrected.

    Although the authors have obviously been very careful with their notations, in particular as concerns indices, I have noticed a few typos

33. **Ll. 177 and 184, $\mathbf{D}_{i;j}$ should be replaced with $\mathbf{D}_{i-j;k-l}$ and $\mathbf{Y}_{i-j;k-l}$ respectively**

    Reply: Corrected, thank you! (l.177, 184)

34. **Eq. (32), (22) above first = sign should rather be (30)**

    Reply: Eq. (32) can be derived either from using Eq. (32) (new version = (29) old verion), (33) (new version = (30) old verion) or (23) (new version = (22) old version). We replaced Eq. (22) by both, Eq. (32) and (33) because each of them refers to one of the innovation cross-covariances used here (see Eq.(35) in new version).

35. **L. 442, $m_{f-1} \rightarrow m_{F-1}$**

    Reply: Corrected, now m_{G-1} after changing index according to comment 5 - reply part B above. (l.480)

And finally, it is the first time I have seen superscripts above equality signs to give reference to previous equations. I think that can be useful, and it is undoubtedly in the present case.

Reply: Thank you for the supporting comment.