Response to Editor (Olivier Talagrand):

Two referees have now sent their reports on the paper, and the open discussion of the paper has been closed. The authors have been asked to post their responses to the referees' comments by 17 Jan 2023.

Referee 1 has let his name known, and is Ricardo Todling. He considers that the work presented in the paper is sound, mathematically meaningful and represents an important contribution to the field. He asks for minor revisions, and gives a number of suggestions to that end, some of which are rather broad, while others bear on editing aspects of the paper (in particular on the English).

Referee 2, who has remained anonymous, is more critical. He writes that he has not fully understood the very logic of the paper. What is the problem, beyond solving an underdetermined system of linear equations by a priori specification of the values of a sufficient number of unknowns ? He also raises the question of the symmetric non-negative character of the variance-covariance matrices produced by the estimation process described in the paper. He asks for major revision of the paper.

As Editor, I suggest to the authors to prepare of course their responses to the referees, and also (if they have not already done so) to prepare a revised version of their paper. This new version must take into account the questions, comments and suggestions of both referees. In case the authors disagree with a particular referee comment, or decide not to follow a particular suggestion, they will have to give explicitly their reasons for that in their responses.

As Editor, I will in any case not take any decision until the authors have responded to the referees' comments. The authors can in the meantime get in touch with me if they wish, either directly or through the public discussion.

Reply: We thank the editor for the suggestions. We prepared a detailed response to each of the referees and prepared a revised version of the manuscript. We hope that we could reply and adopt the manuscript in a sufficient way. In case of remaining questions or suggestions, we are open for further discussions.

<u>**Response to Reviewer 1 - Ricardo Todling:**</u>

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

The present work revisits the problem of estimating relevant statistical information for data assimilation by employing residual-based collocation methods. The work presents a generalization of three-cornered-hat (3CH) and traditional collocation methods establishing precise statements about how many relevant statistics can be inferred from a given number of datasets that include different estimates of sought out quantities. The work also provides for an understanding of what one can expect to estimate given various dependencies among differing datasets. The work is full of insight and provides illustration from idealized settings.

I my view the work is sound, mathematically meaningful and represents an important contribution to the field. I recommend some revision in the text, mostly minor points. I do have a couple of broad comments which are presented below before the minor points.

Main concerns:

A **My first main concern refers to the wide use of the word innovation. Although I understand the main motivation behind the work is data assimilation applications, the framework in the present article is general - it deals with second moment statistics of variables regardless of the context in which these appear. The difference fields appearing in equations such as (4) are what would be better referred to as residuals. I strongly suggest replacement of the word innovation with residual. Indeed, unless commenting on related works truly using innovations (e.g., Tandeo et al. 2020; Todling et al 2022; and others), most of the time the authors can omit either of the words; especially once stated initially that the covariances and cross-covariances dealt with in the work are really residual covariances and cross-covariances.**

<u>Reply:</u> This comment includes 2 points.

1) The replacement of the word "innovation" by "residual", which has been applied everywhere not explicitly referring to data assimilation (incl. Fig. 2-4 and Apx. A). Those sentence has been slightly modified to clarify the relation of the word "innovation" in data assimilation to "residuals" (ll.16-18, new count):

   " A number of approaches to estimate optimal error statistics make use of residuals, i.e. the innovations between observation and background states in observation space (Tandeo et al., 2020), but the error estimation problem remains underdetermined. "

   And under consideration of Minor comment 26 below (ll.327-328, new count):

   " The independence assumption resembles the innovation covariance consistency of data assimilation, were the residual covariance between background and observation datasets - denoted as innovation covariance - ... "

   In the labels of Fig. 2-4, "innovation" was replaced by "residual cov" and in addition the following other labels were extended to be more specific and consistent: "dependency" → "err dependency", "err(*)" → "err cov (*)", and "dep(*)" → "err dep (*)" (see Fig. 1

below, the other figures were modified accordingly, which are Fig. 2-4 in the new version of the manuscript).

2) The suggestion to omit either of the words. With great respect to the reviewer, we believe that it is important to keep the word "residual" (or "innovation") in the text. We see that the word appears quite often, but it the manuscript deals with error and residual statistics at the same time and a clear distinction among them is essential for the understanding.

B **Another issue for me relates to notation. It starts around line 139, when the authors introduce eq. (4). I understand that subscripts such as $i - j$ represent differences (residuals) derived from estimates x_i and x_j for datasets i and j respectively. It is never said that in such case, i must never equal j, as it would not make sense to calculate residuals of a dataset against itself. An alternative notation for the subscripts of Γ would be $i, j; k, l$ - in this, the pairs being use to calculate the difference vectors making up Γ are separated by a semicolon. This notation would also be more consistent with the notation in eq. (5), when the truth is introduced and the matrix represents an error covariance.**

Reply: This comment also includes 2 points.

1) As suggested, a note on non-zero residuals $j \neq i$ and $l \neq k$ was added to the definitions of residual cross-covariances (l.150, new count):

   " Let $\Gamma_{i-j;k-l}$ be the residual cross-covariance matrix between dataset residuals $i - j$ and $k - l$ with $j \neq i$ and $l \neq k$, were ... "

   And similarly in the definitions of residual covariance, dependency, and asymmetry (l.157, l.176, and l.183, new count).

2) Concerning the index-notation of residual statistics, we decided to keep the current notation. In our point of view, the both notations for residuals are consistent with the tilde-notation for error statistics in Eq. (5). However, we find that the suggested comma-notation $\Gamma_{i,j;k,l}$ makes it difficult to distinguish between the two pairs of datasets as commas and semicolons look quite similar, especially for small subscripts. We choose the minus-notation $\Gamma_{i\text{-}j;k\text{-}l}$ because we think that the explicit formulation of the differences used makes the interpretation highly intuitive.

C **I believe that in the considerations in section 3, and specially section 4, a relevant possibility for how possibly to get the precise estimates when dependence exists among the datasets has been overlooked. The others talk a lot about what happens when the dataset are truly independent, or when there is dependence. But never really point out the important case when the dependent contribution in, say, eq. (26) vanishes as a whole. That is when, the datasets are such that**

$$\mathbf{D}_{\tilde{i};\tilde{j}} + \mathbf{D}_{\tilde{k};\tilde{i}} - \mathbf{D}_{\tilde{j};\tilde{k}} = 0$$

The above is at the core of the Todling et al. (2022) findings. That is when the three datasets, (i,j,k) here, are connected in some very special (particular) way, that is, through the DA system, (i.e., these being the analysis, background and observation). I believe the possibility of finding special combination of datasets (for which the above holds) should be discussed in your work. Clearly, datasets that combine in such particular way are rather rare.

Reply: Indeed, the special case of vanishing dependencies is an important point. We already had a short note in Sect. 4.1.3 stating that one dependent contribution might vanish even if all three dependencies are non-zero (compare l.382, new count). We extended this discussion w.r.t. the special case of data assimilation (ll.384-386, new count):
" A special case was observed by Todling et al. (2022) who showed that the estimations of background, observation and analysis errors in a variational data assimilation system become exact if the analysis is optimal. In this particular case, no assumptions on dependencies are required because the optimality of the analysis induces vanishing dependencies. "

In addition, the particular case of variational data assimilation referring to Todling et al. (2022) was also added in the introduction as suggested in Minor comment 7 (compare reply to this comment below).

Minor points:

1. **I wonder about the title a little bit. The work here is very general, I know data assimilation is the primary motivation for the application of the method(s) discussed and the work done in this work. But the fact is that the technique here applies generally, and independently of DA. Perhaps a better title could be: "When do collocated data provide for a closed error estimation problem?"**

   Reply: The words *"in data assimilation"* were removed from the title to express the generality of the work. The title reads now:
   " How far can the statistical error estimation problem be closed by collocated data? "

2. **l. 24: "arises the question if" should read "raises the question whether".**

   Reply: Corrected.

3. **With extreme respect to the authors, I recommend a close revision of the writing itself. I find use of very uncommon English words, which although not incorrect, seem rather usual, e.g., exemplary, approximative; there are also a number of articles, and other wording in the paper that could benefit from some attention. I try to point out some of these in what follows, but I only show so much. I can anticipate that most of the time a word like "exemplary" is better read as "example", and "approximative" as "approximate".**

   Reply: We corrected the two examples given here as well as all words given in the other comments below. In addition, we:

   - corrected some typos ("accuracies", "structural" in Sect. 1, "triangle" in Sect. 5)
   - did some corrections to achieve a consistent hyphen ("-") notation for words with two parts (e.g. "state-vector", "grid-points", "element-wise") and in listing of words (e.g. "covariance- and cross-covariance matrices")
   - we also removed repeating abbreviations "Eq." and "Sect." in listings of several equation or section numbers (e.g. "Eq. (36), (37), or (38)", l.410, new count)

4. **l. 30: "since decades" should read "for decades".**

   Reply: Corrected.

5. **l.30: "recently be exploited" should read "recently been exploited".**

   Reply: Corrected.

6. **l. 31: The works of Nielsen et al. (2022) and Todling et al. (2022) were done concurrently, basically with either being unaware of the details of the other. I believe your statement here would more fairly read "Nielsen et al. (2022) and Todling et al. (2022) were the first to independently use the generalized ...".**

   Reply: Modified accordingly.

7. **ll 33-35: I think the authors need to rephrase what comes after "However". A better sentence would perhaps be: "...framework. Indeed, Todling et al. (2022) shows that when the corners of G3CH are identified with the observation, background and analysis of variational assimilation procedures, only under the assumptions of optimality does the method obtains closed estimates for the three corners; in general, the problem cannot be closed." Notice this comment goes along comment C made above.**

   Reply: Rephrased accordingly, the sentence reads now (ll.34-36, new count):
   " They show that when the corners of G3CH are identified with the observation, background and analysis of variational assimilation procedures, this particular error estimation problem can only be closed under the assumption of optimally. "

8. **The "dot" notation used in eq. (4), and many others, has not been introduced. I authors should state that**

   $$x \cdot y = xy^T$$

   Reply: Actually, $x_i(p)$ denotes one element of the dataset vector, thus the dot denotes a scalar multiplication with no need for a transposed. We removed the dot for all multiplications of vector-elements (Eq.(4)-(7),(19),(21)) and added the following note for clarification (l.156, new count):
   " Note that $x_i(p)$ is a scalar element of the dataset vector. "

9. **Eqs. (6) and (7) do not require the term exposed in their second equalities.**

   Reply: That's right. However, the authors believe that an explicit formulation of these quantities helps the reader when re-checking the definitions without reading the whole section.

10. **Eq (6): I confess the notation of using superscripts in the equality signs in various equations is new to me. I have mixed feelings about it, but regardless of my feelings, the authors should explain what these are after they first appear in eqs. (6) and (7). That is, somewhere it should be stated that "superscripts and subscripts in the equality signs indicate what other equations were used to arrive at the given result".**

    Reply: Although not being widely used, superscripts above equal signs have been used previously to indicate other equations used (e.g. Vogel and Elbern, 2021, GMD). Regarding the number of other equations that were used in some derivations in this manuscript (e.g. Eq. (34)) we decided to use this notation to help the reader following the derivations in a compressed way that does not affect the reading flow significantly. But we agree that the notation has to be introduced and thus removed the superscripts in Sect. 2 (in Eq. (2) and Eq. (3)) and added the following sentence were it first appears in Sect. 3 (l.161, new count):

" ... were numbers in parenthesis above an equal sign indicate other equations that were used to retrieve the right hand side. "

Concerning the use of indices of these superscripts, we removed these indices whenever not necessary (Eq. (27), (29), (45), (46), (47), (48), and (53)) and only kept them were the same equations were applied to different datasets in order to avoid potential confusion. In these cases, the descriptions of the equations were extended. For Eq. (25) (ll.245-247, new count):
" By combining the formulations of three residuals $\Gamma_{i\text{-}j}$, $\Gamma_{j\text{-}k}$, and $\Gamma_{k\text{-}i}$ between the same three datasets i, j, and k and expressing each using Eq. (20), a single error covariance can be eliminated: ... "
and for Eq. (30) (l.283, new count):
" Two of the error cross-covariances in Eq. (29) can be rewritten by applying Eq. (29) to the error covariance of dataset j: ... "

We also added a description were these indices were used for the first time (ll.252-254, new count):
" ... were the indication of used equations above the equal signs are extended by indices which denote to which datasets this equation has been applied. For example, " $\overset{(20)_{ki}}{=}$ " indicates that the relation in Eq. (20) was applied to datasets k and i to achieve the right hand side. "

The meaning of subscripts of equal signs (indicating the assumptions used in this relation) was already described in the manuscript were there first appear (l.331. l.339, and l.414, new count), were l.331 slightly extended to:
" ... were " $\underset{\{in\}}{\approx}$ " indicates the assumption of independence between the two datasets, i.e. $X_{\tilde{i};\tilde{j}} = 0$. "

And similarly l.414:
" ... were " $\underset{\{inI\}}{\approx}$ " indicates the assumption of independence to the reference dataset, i.e. $X_{\tilde{i};\widetilde{ref(i)}} = 0$. "

11. **l. 47: "since decades" should read "for decades".**

    Reply: Corrected.

12. **l. 53: "additionally" should read "additional".**

    Reply: Corrected.

13. **l. 70: word "approximative" would better read "approximate". The word approximative appears numerous times, I believe all instances would read better as "approximate" instead.**

    Reply: Replaced everywhere were it appeared.

14. **l. 76: word "exemplary" should be removed in this case - without loss of clarity.**

    Reply: Removed.

15. **l. 79: "requiring the knowledge" would better read "requiring knowledge".**

    Reply: Corrected.

16. **ll. 84-85: "analyses or any" would better read "analysis and any".**

    Reply: "or" replaced by "and". "analyses" was kept in its plural form to remain consistent with the other listed items (l.87, new count).

17. **l. 138: there needs to be an explanation (definition) for the meaning of the subscript notation with the standing up bar, as in $i|r$, i given r? Why do you need this notation here when it is not used anywhere else in the article?**

    Reply: We agree that this notation might be confusing and is not important for the rest of the manuscript. We removed the explicit indication of the realization in the definition of the state vectors and deleted the equation in the 2nd sentence, which now reads (ll.147-148, new count):
    " Suppose I datasets, each containing R realizations of spatio-temporally collocated state vectors x_i $\forall$ $i \in [1, I]$. Without loss of generality, the following formulation uses unbiased state vectors with zero mean. "

    Instead, we added a note on the meaning of the overbar w.r.t. realizations before Eq.(4):
    " ... were each element (p, q) is given by the expectation over all realizations: ... "

    And after Eq.(5):
    " ... and the overbar denotes the expectation over all R realizations. "

18. **Eq. (4): I find it somewhat unnecessary to have the notation include the points (p; q) explicitly. Given that x is a vector quantity the (p; q) indexes can be implicitly understood. In fact, most of your eqs. do not carry them.**

    Reply: Although most equations refer to complete matrices, we believe that the complexity of the terms itself - especially when carrying multiple indices - justify the explicit definition of a single element of the matrix. While it might be obvious to some readers, the formulation of a single element avoids misunderstanding e.g. in Eq. (21).

19. **ll. 173-174: This sentence should be moved to the definition statements made around eq. (4).**

    Reply: The sentence was moved to the end of the introduction paragraph of this section, before Eq.(4) (now ll.48-49, new count).

20. **ll. 190-193: This would better read: "Thus, the covariance of any two datasets consists of the sum of the independent covariances associated with each dataset minus the error dependency covariance; this latter corresponding to the sum of the error covariances associated with each dataset, eq. (16)."**

    Reply: Rephrased based on the reviewer's suggestion. The sentence now reads (ll.201-203, new count):
    " Thus, the residual covariance of any dataset pair consists of (i) the independent residual associated with sum of the error covariances of each dataset, minus (ii) the error dependency corresponding to the sum of their error cross-covariances. "

21. **l. 240: "formulated as sum" should read "formulated as a sum".**

    Reply: Corrected.

22. **paragr. ll. 243-247: you might want to add here that all the works mentioned in this paragraph associate what the authors call "dependent contribution" with the cross-covariances of the random errors.**

    Reply: The following sentence was added (ll.264-265, new count):
    " Note that in the literature, the dependent contribution in Eq. (26) is denoted as cross-covariances between the errors. "

23. **l. 243: please replace "by Eq. (26)" with "in Eq. (26)".**

    Reply: Done.

24. **ll. 260-261: There are lots of instances of the word "formulation" in these two sentences; the author might want to work on the text.**

    Reply: Rephrased. The sentences read now (l.278-279, new count):
    " The scalar formulation of Eq. (29) was previously given in Zwieback et al. (2012). Similarly to Eq. (26) from residual covariances, the number of formulations .... "

25. **ll. 278-279: This sentence is very confusing. I think I understand what the authors mean, but I suggest rephrasing.**

    Reply: Rephrased. The sentence now reads (ll.296-297, new count):
    " Note that the third dataset i on the right hand side of Eq. (33) can be any other dataset ($i \neq j, i \neq l$). Thus for any $I > 2$, there are $I - 2$ formulations of each error cross-covariance $\mathbf{X}_{\tilde{j};\tilde{l}}$, which are all equivalent in the exact formulation. "

26. **ll. 306-308: I believe the authors want to say that the "*independence* assumption *resembles* the innovation consistency *statement* of data assimilation, where the innovation covariance..." - notice that here, this is one of those places where the word *innovation* can and should be used.**

    Reply: Replaced accordingly; also at all other locations in the manuscript were "independent assumption" appeared. As suggested by the reviewer, the term innovation was kept here, with slight modification as described in reply to Main concern A.1.

27. **l. 314: word "neglectable" should read "negligible".**

    Reply: Replaced; also at all other locations in the manuscript were it appeared.

28. **l. 325: word "between" should be replaced with "among". Please notice there are other instances of "between all three" that should be revised accordingly.**

    Reply: Replaced whenever referring to three or multiple datasets.

29. **l. 337: "a error" should read "an error"**

    Reply: Corrected.

30. **l. 339: "allow a comparison" should read "allow for a comparison".**

    Reply: Corrected.

31. **l. 396: " Eq. (32) is follows" should read " Eq. (32) follows".**

    Reply: Typo. Replaced by "it follows" (l.418, new count).

32. **l. 418: spell: "beeing".**

    Reply: Corrected.

33. **l. 461: "An discussion" should read "A discussion".**

    Reply: Replaced.

34. **l. 464: "provides an exemplary demonstration" would better read "provides some demonstration".**

    Reply: Replaced with "illustrates" to be consistent with the comment on l.495 below (l.488, new count).

35. **l. 485: word "calculated" is not needed.**

    Reply: Removed.

36. **l. 495: "exemplary demonstrated" should better read "illustrated".**

    Reply: Replaced.

37. **l. 511: "does only affect" should read "affect only".**

    Reply: Replaced.

38. **Fig. 2a: why are the errors (bottom row) so diagonally dominant? Shouldn't these bottom panels be more like random patterns everywhere? Why aren't the errors in the diagonal of the order of the off-diagonal terms?**

    Reply: We are not entirely sure if we understand this comment the right way, we see two possible interpretations which we will both consider in the the following reply.

    (a) Looking at the bottom rows, the upper left part shows the absolute differences between true and estimated statistics. In the case of Fig. 2a, all assumptions are sufficiently fulfilled and the estimated statistics becomes equal to the true statistics. Indeed, what remains is some minor random noise. The reviewer might refer to the gray diagonal bar which separates the triangular matrices and has nothing to do with the fields itself. We see that the existence of the gray line might lead to miss-interpretation in the 3rd row and removed it accordingly. Additionally, we filled the lower right triangle with gray stripes to indicate that there is no data (rather than zero-values) shown in this part (see Fig. 1, the other figures were modified accordingly, which are Fig. 2-4 in the new version of the manuscript). The description of the plots in the manuscript was also extended for clarification (ll.500-502):

    " Because all matrices involved are symmetric, it is sufficient to show only one half of each matrix. The two matrices are separated by a thick gray diagonal bar and shifted off-diagonal so that diagonal variances are right above/below the gray bar, respectively. "

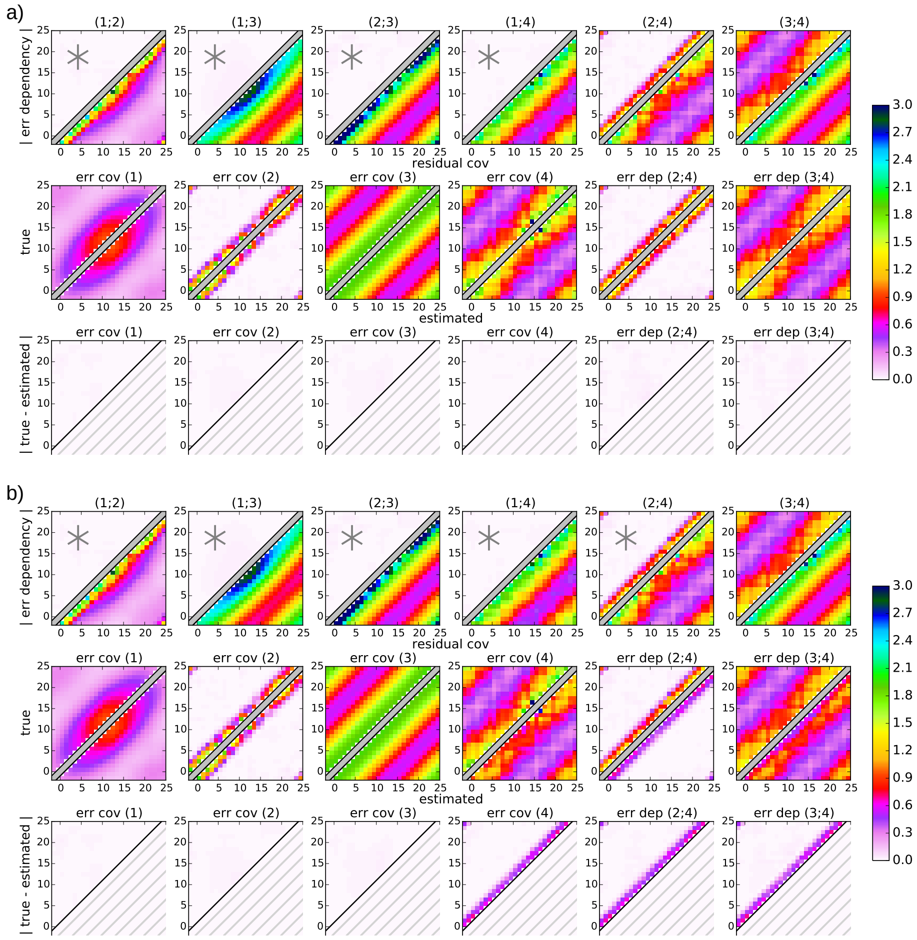

Figure 1: Covariance matrices for 4 datasets ($I = 4$) with true dependencies of datasets (2;4) and (3;4). Datasets (1;2;3) build the basic triangle. Dataset 4 is estimated (a) from its reference dataset 1 ("sequential estimation") and (b) from an additional independent triangle (1;2;4) ("triangular estimation").

(b) Otherwise, the reviewer might refer to the fact that differences between true and estimated statistics are only visible close to the diagonal in Fig. 2b,3,4. This is a result of the shape of error dependencies that were neglected in the estimation (upper part of 1st row, indicated by a gray asterisk) which are created to be proportional to the referring residual statistics (lower part of 1st row). Only the smaller amplitude might make the uncertainty appear more diagonal-dominant than the residual covariance, if this is what the reviewer is referring to.

39. **Figs. 2b and 3: why are the errors (bottom row plots) so asymmetrically dominant?**

    Reply: This comment is closely related to the previous one. The gray diagonal bar together with the upper-triangular field and the missing data in the lower triangle might have appeared as asymmetric field. See reply to previous comment 38, the applied corrections should also clarify this point.

40. **l. 515: "it's" should read "its".**

    Reply: Corrected.

41. **l. 516: "requirement of assumptions" should better read simply "assumptions".**

Reply: Replaced by (l.543, new count):
" assumptions and requirements "

42. **l. 522: "and 4.2 and is" should better read "and 4.2 is".**

    Reply: Removed.

43. **l. 523: please spell out "Apx".**

    Reply: Done; also were it appears elsewhere in the manuscript.

44. **l. 530: "solution of the problem" reads better as "solution to the problem".**

    Reply: Replaced.

45. **l. 533: word "whoever" should be removed.**

    Reply: Removed.

46. **l. 533: "came up with a too strong requirement" would better read "came up with an unnecessarily strong requirement that ldots"**

    Reply: Replaced.

47. **l. 539: "estimates" should be in the singular.**

    Reply: Corrected.

48. **l. 575: duplicate "of the".**

    Reply: Removed.

49. **p. 24, conclusions: in regards to your last two paragraphs, and the generality of the method as you propose here, can you comment on the viability of the method to be used for, say, deriving estimates of background errors by using a combination of background fields from multiple DA systems. For example, suppose we collect 6-hour forecasts from IFS, GFS, CMC, GMAO, US Navy, etc - there is some dependency among all these datasets since for most part the background filds (short-range forecasts) are based on the assimilation of similar observations in all these systems - do you think your method would be able to infer reliable and perhaps better forecast error estimates than what we typically get from the NMC or ensemble methods? The same question can be made wrt analysis errors. Can you comment on this - if not in the paper, at least here to this reviewer.**

    Reply: Thank you for this thoughtful suggestion. We added the following paragraph in the conclusions (ll.622-626, new count):
    " An important application of the presented method is expected to be numerical weather prediction (NWP) were short-term forecasts from multiple national centers can be used to estimate error statistics required for data assimilation. In contrast to previous statistical methods, potential dependencies among the forecasts, i.e. due to the assimilation of similar observations, can be considered in the error

estimation and even explicitly quantified. Future work will show how this statistical approach compares to state-of-the-art background error estimates based on computation-expensive Monte-Carlo- or ensemble-methods. "

We decided to keep the discussion of how the presented method might compare to state-of-the-art estimates rather short in the manuscript. This will be left for a follow-up investigation were the method is actually applied to real geophysical datasets. The authors expect the main advantage of the proposed method to be its very low computation effort given that a large number of overlapping datasets are already available (eg. global gridded opertational forecasts from different weather centers worldwide which only need to be interpolated to a common grid). Another advantage that is also mentioned in the manuscript is the explicit estimation of error dependencies (or cross-covariances), which however require the development of novel data assimilation schemes.

Despite the need for collocated datasets, the main disadvantage of this method lies in its attempt to estimate statistical covariances only. The need for a large set of realizations which sample the same truth is expected to be a limitation for many real applications. This aspect will also be further discussed and different solutions will be proposed in the upcoming work.

Response to Reviewer 2:
We thank the reviewer for the insightful remarks and hope that we could reply and adopt the manuscript in a sufficient way.

One major question in data assimilation is to determine the statistics of the errors affecting the data to be assimilated. It is those statistics that define the weights to be given to the data in the assimilation. However, they can never be fully determined without external hypotheses, i. e. hypotheses that cannot be objectively validated on the basis of the data alone.

The authors present and discuss an approach that is appropriate for the situations in which a number of sets of collocated data are available. They consider only second-order statistical moments (first-order moments, i. e. biases, are also required, but their identification is an independent problem). Covariances and cross-covariances of the differences ('innovations') between those different sets of data are known from the data, and are linearly related to the statistics of the underlying data errors. By appropriate a priori specification of a number of those data errors statistics (the external hypotheses), all, or part, of the remaining error statistics are solution of a system of linear matrix equations. That approach, which originated from the so-called three cornered hat (3CH) method, has been used in a number of applications, but not much so far in assimilation of geophysical data.

Given I sets of collocated data, the unknowns (covariances and cross-covariances of data errors) are in number $U_I = (1/2)I(I+1)$ (Eq. 2 of the paper). Concerning the innovations, their second order moments are not independent, and they are combinations of only $N_I = (1/2)I(I-1)$ of them (l. 95). This leads to a linear system of N_I matrix equations with U_I unknowns (that system is basically expressed, although in what is to me a cursory passing remark, by Eq. 22). The degree of underdeterminacy of the system is $U_I - N_I = I$. The view that is suffices to choose a priori I of the unknowns to close algebraically the system is correct for $I = 3$, but not necessarily for larger values (at least if, as the authors want, no error covariance is specified a priori). The purpose of the authors, in addition to stating precisely and discussing the problem, is to determine minimal conditions for its solution (... what are the minimal and optimal conditions to solve the problem?, l. 65). They also present numerical results obtained from synthetic data.

The article is instructive, and certainly contains material that is worth publishing. But it needs in my opinion substantial improvement.

1. **My main comment is that I have found it very difficult to understand the very logic of the paper (and I am actually still not even sure I have fully understood). A succinct analysis shows that, for $I > 1$, system (22) (strictly speaking, a system of N_I equations which is equivalent to 22) is of rank N_I, which shows that by appropriately choosing I of the unknown error covariances and cross-covariances, one can obtain the values of all the other unknowns. My understanding is that the authors show that these I a priori chosen error covariances and cross-covariances cannot be chosen arbitrarily, and that there are constraints in that choice (especially in the case considered by the authors, in which only cross-covariances are to be chosen a priori). If it is so, I think it must be stated more explicitly.**

   Reply1: The reviewer is right in his interpretation. We intended to describe this important aspect in Sect. 2, but we see now that it was not formulated clearly enough. We reformulated the referring paragraph, which now reads (ll.103-105, new count):

" The set of error statistics to be estimated can generally be chosen according to the specific application, but it will be shown that there are some constraints. Based on the mathematical theory provided in the following sections, the actual minimal conditions to solve the problem will be discussed in Sect. 6.1. "

We also added a new paragraph at the end of this section (ll.128-130):

" Note that almost all numbers presented above apply to the general case were any combination of error covariances and cross-covariances may be given or assumed. While the interpretation of the numbers I, N_I and U_I remains the same in all cases, the only difference is the interpretation of DI which is less meaningful when also error covariances are assumed. "

We ensured that the actual formulation of the minimal conditions (Sect. 6.1) already includes the information that those are formulated for the common case that only cross-covariances are assumed (see eg. l.553). Additionally, we added a note that similar conditions holding for other cases at the end of the subsection (ll.569-570, new count):

" Note that similar conditions can be derived for cases were also error covariances are given or assumed, which is not part of this paper. "

2. **Subsection 6.1 (Minimal conditions) contains what I understand are the authors' main conclusions. That Subsection states two conditions (ll. 527-529) that are presented as the minimal conditions ensuring existence and uniqueness of the solution of system (22) (at least, that is my understanding)**

   (i) all three error dependencies between one triple of datasets are needed (this triple of independent datasets is called "basic triangle")

   (ii) at least one error dependency of each additional dataset to any prior datasets is needed

   I is not clear to me whether these two conditions are mathematically exact (if yes, explain more clearly where they are proven in the paper, or give a reference ; if not, say clearly they are only reasonable conjectures).

   Reply2: Based on the mathematical derivations in Sect. 3 and 4, the two minimal conditions are logical conclusions that are valid for all number of datasets. They provide the necessary conditions for the existence of a solution; which is demonstrated by giving the explicit formulations of error statistics in Sect. 4.1.1, 4.2.1 and 4.2.2. The uniqueness of this solution is achieved when - and exactly when - the required assumptions are accurate; i.e. the assumed error cross-covariances and dependencies vanish (compare equations in Sect. 4.1.3, 4.2.3 and 4.2.4). A refering statement was added in the manuscript (ll.557-561, new count):

   " These two requirements are a logical summary of the mathematical derivations in Sect. 3 and 4 and are valid for all number of datasets $I \leq 3$. They provide the necessary conditions for the existence of a solution under the given assumptions (compare Sect. 4.1.1, 4.2.1, and 4.2.2)). Optimality and uniqueness of this solution w.r.t. different formulations and setups are achieved when - and exactly when - the required assumptions are accurate (i.e. vanishing uncertainties of assumed error statistics in Sect. 4.1.2, 4.1.3, 4.2.3, and 4.2.4). "

   A rigorous mathematical proof of these conditions would require another level of math in this paper, which is mainly tailored to a theoretical-geophysical rather than a mathematical audience. Therefore, we decided to not include a rigorous proof in this paper.

Note that the given minimal conditions apply for setups *"were all error covariances and some error dependencies (or cross-covariances) are estimated"* (l.553, new count).

3. **I find that Sections 3 and 4, although they boil down to elementary algebraic manipulations, are intricate and difficult to follow.**

   a. **Eq. (22) expresses the basic links between innovation and error statistics (denoted respectively Γ and X). Although algebraically obvious, it is the crux of the method, and should be stressed more strongly as such.**

   Reply3a: We agree with this comment and thank the reviewer for pointing this out. We pointed out the importance of Sect. 3.2 and specifically Eq. (20) at several locations in the manuscript:

   - In the description of the content of the study in Sect. 1 (ll.69-72, new count):
     " ... the mathematical formulation for non-scalar error matrices is derived in Sect. 3 and Sect. 4, respectively. The derivation is based on the formulation of residual statistics as function of error statistics which is introduced in Sect. 3.2. While the exact formulations of error statistics in Sect. 3.3 remain underdetermined in real applications ... "
   - In the description of the general framework in Sect. 2 (ll.96-97, new count):
     " The main idea now is to express the known residual statistics as function of unknown error statistics (Sect. 3.2) and combine these equations to eliminate single error statistics (Sect. 3.3, Sect. 4). "
   - In the introduction of Sect. 3 (ll.137-138, new count):
     " The expression of residual statistics as function of error covariances and cross-covariances in Sect. 3.2 provides the basis for the subsequent mathematical theory. "
   - In the introduction of Sect. 3.3 (ll.239-240, new count):
     " These formulations are based on the relations between residual and error statistics in Eq. (20) and Eq. (22). "
   - And in the section itself along with the discussion of the equation (ll.216-217, new count):
     " This formulation of residual statistics as function of error statistics provides the basis for the complete theoretical derivation of error estimates in this study. "

   b. **The derivation of Eq. 23 (ll. 211-213) is strange, since it suggests (l. 211) that one must go through the error statistics X to obtain the equation, while the latter expresses necessarily links between the innovation statistics Gamma, and can be easily be proved directly.**

   Reply3b: We see that this formulation is confusing. We rearranged the equation that the individual steps to show the equivalence between innovation statistics in a clearer way (Eq. (23), new count).

   The decision to use error statistics to show the equivalence was made upon its context in the work which is based on the relation between innovation and error statistics (also pointed out by the reviewer e.g. in his comment 3a above).

   c. **Eq. (34) is also strange in that in purports to show the 'equivalence' between two expressions for the error dependencies D. Those two expressions are basically obtained from Eq. (22), and the reader would think they must necessarily be the same. I presume the authors want to stress that inappropriate choice of the a priori chosen error cross-covariances can lead to inconsistencies.**

But, rather than demonstrating consistency, it would be preferable to show an explicit example of inconsistency. Actually, my understanding is that Eqs (39-40) precisely show an example of inconsistency. If I am mistaken about the significance of Eq. (34), say more explicitly what that significance is.

Reply3c: Indeed, the equivalence of the two expressions is not surprising. The formulation based on innovation covariances originates form Eq. (20) which is a special case of the basic equation for cross-covariances Eq. (22) (compare also Reply3a). However, the authors decided to show this equivalence to explicitly demonstrate the consistency of the two estimates of error dependencies. While this was obvious to the reviewer, we think that it will be useful to other readers. Given the fact that previous literature considers only one of the two formulations, either based on residual covariances or cross-covariances, this relation might not be clear to everyone. At the same time, the equivalence is an important result of the work, which refers to the last paragraph of the reviewers comment 3 (see below). We added the following comment to the description of Eq. (34) (ll.304-306, new count):

" This consistency applies to the exact formulations of all symmetric error statistics (error covariances and dependencies) and results directly form the fact that the basic formulation of residual covariances in Eq. (20) is a special case of the formulation of residual cross-covariances in Eq. (22). "

d. **The authors, for some unspecified reason, consider only the 'error-dependencies', i.e. the symmetric part of the error cross-correlations matrices (Eq. 20), and ignore the antisymmetric part. Why so ?**

Reply3d: The reference to manuscript is not entirely clear. In Sect.3 and 4, the authors carefully verified that asymmetric cross-covariance matrices and asymmetry matrices are considered wherever possible and if not, a related statement were made (eg. ll.254, ll.403, old version). What remains are the experiments in Sect. 5 which indeed only discuss the symmetric statistics (error covariances and dependencies) and not asymmetric error cross-covariances. This was the author's choice in order to restrict the section to it's main purpose of demonstrating the ability to retrieve error covariances as well as dependencies. Actually, the presented experiments were created with symmetric statistics which only enables the compressed visualization in Fig. 2-4 (showing only half of each matrix). The demonstration could be extended to explicitly show asymmetric error statistics, but we believe that showing dependencies is more convenient for this purpose (shorter and more intuitive); especially under consideration that the manuscript is already quite long. We added a referring sentence in the description of the experiments (ll.496-498, new count):

" Similar results would be obtained from estimations from cross-covariances in Algorithm A2, but this short illustration is restricted to a general demonstration using symmetric statistics only. "

and in the explanation of the plots (ll.500-501, new count):

" Because all matrices involved are symmetric, it is sufficient to show only one half of each matrix. "

e. **It is not clearly said why the number of independent innovation covariances and cross-covariances is equal to $N_I = (1/2)I(I-1)$ (that is rather simple, but must be said more clearly). The mutual dependence between those quantities is expressed by Eq. 23, the significance of which (in addition to my remark b above) should be stressed more strongly.**

Reply3e: We clarified the explanation for the number of independent innovation statistics (now denoted as "residual statistics" based on Reviewer1, ll.97-100, new count):

" Because of $j \neq i$ for residuals, each of the I datasets can be combined with all other $I-1$ datasets. As residual statistics also do not change with the order of datasets in the residual (see

Sect. 3.1), the number of known statistics of the system is also given by N_I as defined in Eq. (1). "

A note on the significance of Eq. (23) - beeing the main equation in this Sect.3.2.3 - was added in the description of the general framework Sect. 2 (ll.101-101, new count):
" It will be shown in Sect. 3.2.2 that residual cross-covariances contain generally the same information as residual covariances; thus the N_I residual statistics can be given in form of residual covariances or cross-covariances. "

These are only examples of places that can cause confusion in the mind of a reader who is a newcomer to the approach described in the paper, as elementary as that approach may fundamentally be. I think Sections 3 and 4 could be rewritten in a clearer and more concise way, with more stress on the logic of the approach and on the two fundamental aspects upon which it is based. First, that the observed innovation covariances and crosscovariances are redundant. Second, the basic link between between the innovation and errors covariances and cross-covariances, expressed by Eq. (22) (or any other equivalent equation for that matter).

Reply3: We hope that the corrections applied w.r.t.comment 3a-e already contribute significantly to the clarification. Firstly, a reference to the equivalence of residual covariances and cross-covariances was made in the formulation of the general framework (Sect. 2) for comment 3e. Secondly, the importance of the link between residual (innovation) and error statistics, which was emphasized at several locations in the manuscript for comment 3a - including the introduction (Sect. 3) and general framework (Sect. 2) and the introduction of the theoretical part (Sect. 3 and 3.3). Thought these changes, the formulation of the general framework in Sect. 2 now introduces explicitly the two fundamental aspects, puts them into context of the framework and provides references to the respective sections in the theory.

In addition to the modifications above, the list of new aspects in the introduction of Sect. 3 was extended w.r.t. the fundamental aspects suggested by the reviewer (ll.142-145, new count):
" This first part of the mathematical theory includes the following new elements: (i) the separation of cross-statistics into a symmetric error dependency and an error asymmetry (Sect. 3.1), (ii) the general formulation of residual statistics as function of error statistics (Sect. 3.2.1 and 3.2.2), (iii) the demonstration of equivalence between residual covariances and cross-covariances (Sect. 3.2.3), (iv) the general formulation of exact relations between residual- and error statistics (Sect. 3.3). "

And the list of Sect. 4 was modified accordingly (ll.313-316, new count):
" In addition to the optimal extension to more than three datasets, this second part of the mathematical theory includes the following new elements: (i) the analysis of differences from residual covariance and cross-covariance estimates (Sect. 4.1.2), (ii) the determination of uncertainties caused by assumed error statistics (Sect. 4.1.3 and 4.2.3), and (iii) the comparison of the approximation from three- ("triangular estimation") and more ("sequential estimation") datasets (Sect. 4.2.4). "

4. **And, for a final (but I think important) comment, any algebraic solution to system (22) will not be acceptable in then present context. It must also define a proper (symmetric non negative) global error covariance matrix (in particular, the estimated error covariance matrices $\mathbf{C}_i$ of the various individual datasets, in addition to being symmetric, must be nonnegative). The authors hardly mention this point. Do the conditions (i-ii) stated in subsection (6.1) lead to a proper global covariance matrix ? Since system (22) expresses necessary conditions between error and innovation variances and covariances, I presume that if the a priori specified variances and cross-covariances are compatible with a globally symmetric non-negative**

matrix that is itself compatible with the $\Gamma_{i,j;k,l}$'s (Eq. 22), the estimated variances and cross-covariances will also be. I do not ask the authors to necessarily give a full answer to that question, but it should be clearly mentioned and at least briefly discussed. In particular, if the authors do not have a full answer to that question, it should clearly stated as remaining an open question.

Reply4: We agree with the reviewer that this is an important point to consider when it comes to the application to real data. The equations in Sect. 3-4 and the minimal conditions in Sect. 6.1 do not ensure positive definiteness of the error covariance matrix.

There is already a comment on this in the original version of the manuscript in the formulation of uncertainties for 3 datasets (Sect. 4.1.3, ll.391-394, new count):

" *Thus, the estimated error covariance matrices might not be positive definite if the independent assumption between three datasets is not fulfilled. This phenomena was also described and demonstrated by Sjoberg et al. (2021) for scalar problems. However, the generalization to covariances matrices is expected to increase the occurrence of negative values were correlations between two entries of the state are low, thus relative differences and sampling errors become large.* "

As well as a general comment in the formulation of uncertainties for multiple datasets (Sect. 4.1.3, ll.484-486, new count):

" *Note that the absolute uncertainties presented here only account for uncertainties due to the underlying assumptions on error cross-statistics and not due to imperfect residual statistics occurring e.g. from finite sampling. An discussion of those effects for scalar problems can be found in Sjoberg et al. (2021).* "

Specifically, negative values appear in the estimates for the estimation with a large neglected dependency in the basic triangle (manuscript Fig. 4). Here, we added a note on the appearance of negative values (ll.535-537, new count):

" *This particular setup also demonstrates that uncertainties due to neglected dependencies can become larger than the actual true statistics (here e.g. in the error dependency of datasets* $(2;4)$ *for both estimation methods) which creates negative values in the estimate.* "

In addition, we added a short note in the discussion section when discussing the application to real data (ll.626-628, new count):

" *While the presented method ensures symmetry of error covariances, positive definiteness might not be fulfilled in real applications due to inaccurate assumptions or sampling uncertainties.* "

It may that the response to some of the questions I raise above is available in the literature, in particular in the literature the authors mention. If so, please give precise references.

I would have a number of other comments, bearing on both scientific and editing aspects of the paper, but they are of lesser importance, and I will wait for a possible revised version for mentioning them.

Reply: We performed several detailed corrections concerning wording, notation and scientific content following the suggestions of the other reviewer. If there are remaining comments in the new version of the manuscript, we are happy to implement them accordingly.