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.