Response to Second Round of Comments from Reviewer 1 for Assimilating Compact Phase Space Retrievals of Atmospheric Composition with WRF-Chem/DART: A Regional Chemical Transport/Ensemble Kalman Filter Data Assimilation System

1. ***In 10, however, the term nU_0^T E_m nU_0 = nPhi nSigma nPsi should be replaced with nU_0^T E_m nU_0 = nPhi nSigma nPhi^T.***

   This typographical error has been corrected.

2. ***My comment about applying an eigenvector decomposition to E_m rather than to U_0^T E_m U_0 is still unanswered. I think the authors should at least note in the paper that the latter option is also possible.***

   If Reviewer 1 is referring to the comment that "transforming U_0^T E_m U_0 instead, at a later stage, is an unnecessary complication" we do not agree. First, when the retrieval equation is written as Eq. (3) in the paper, the left non-zero singular vectors of the averaging kernel are a basis for the range of the averaging kernel (the space of quasi-optimal retrievals (QORs)). The left non-zero singular vectors of the observation error covariance are not necessarily a basis for that space because the observation error covariance is a function of the product of the *a priori* retrieval error covariance and the averaging kernel. Second, if we had reversed the order of our transforms, the final form of the transformed observation error covariance would have been nondiagonal. Ending with a diagonalized observation error covariance was necessary to our goal of facilitating use of modern sequential ensemble Kalman filter data assimilation algorithms.

3. ***My understanding of the transformation as in Eq. 24 of Migliorini et al., 2008 is that it is not based, as the authors say in their replies at (iv), on an eigenvalue decomposition of the ensemble forecast error covariance but rather of H' P H'^T, where P can be the ensemble forecast error covariance. The eigenvalues of H' P H'^T are signal-to noise ratios, so that a truncation according to eigenvalues of H' P H'^T is based on the measurements' signal to noise ratio.***

   We agree with this statement.

4. ***It is of course also possible to truncate according to the singular values of H', which in Migliorini et al. 2008 is given by H' = E_m^{-1/2} A, where E_m is the measurement error covariance and A the averaging kernel matrix. The truncation is then applied by left-multiplying by U^T, the left singular vectors of H', which are identical to the eigenvectors of H' H'^T, i.e. identical to the eigenvectors of the covariance of H' x^t when P = I (this is why this procedure is a particular case of the more general eigenvector decomposition of H' P H'^T).***

   We agree with this analysis but do not agree with the conclusion that our method is a special case of Migliorini et al. (2008). Our method does not depend on the singular vectors of $\boldsymbol{H'}, \boldsymbol{H'H'^T}$, or $\boldsymbol{H'PH'^T}$ so the observation that the singular vectors of $\boldsymbol{H'}$ are the same as the singular vectors of $\boldsymbol{H'H'^T}$ does not prove the methods are equivalent. In our method, the QOR profile is first transformed by the left non-zero singular vectors of the averaging kernel and then transformed by the left non-zero singular vectors of the transformed observation error covariance. In Migliorini et al. (2008), the QOR profile is first transformed by the left non-zero singular vectors of the observation error covariance and then transformed by the left non-zero singular vectors of the transformed forecast error variance in observation space. When $\boldsymbol{P} = \boldsymbol{I}$ their second transform is based on the left non-zero singular vectors of the transformed averaging kernel. Those singular vectors are a basis for the space of transformed QORs, but as explained in our response to Comment 1 it does not appear that the left non-zero singular vectors of the observation error covariance are a basis for the space of untransformed QORs. If not, Migliorini et al. (2008)'s first transform effects a truncation of the QOR profile and the methods are not equivalent.

5. ***Your approach is to decompose nU_0^T E_m nU_0 to find its square root (rather than the square root of E_m) and then to project according to the left singular vectors of A rather than the left singular vectors of H'. The two approaches are, if not mathematically identical, practically equivalent. This should be noted in your paper.***

To avoid confusion, our approach is to first project according to the left non-zero singular vectors of $\boldsymbol{A}$ and then decompose $\boldsymbol{U}_0^T \boldsymbol{E}_m \boldsymbol{U}_0$ to find its square root. This comment seems to have that order reversed.

We do not agree that our method is "mathematically identical [or] practically equivalent" to that of Migliorini et al. (2008). As explained in our responses to Comments 2 and 4, it appears that the left non-zero singular vectors of the observation error covariance are not a basis for the space of QORs, and therefore the methods are not equivalent. Additionally, the methods are different because our observation reduction strategy is based on data compression, and Migliorini et al. (2008)'s reduction strategy is based on data rejection. When $\boldsymbol{P} = \boldsymbol{I}$ and the left non-zero singular vectors of the observation error covariance are a basis for the space of QORs, we agree that the Migliorini et al. (2008) rejection criteria (rejecting modes with singular values less than one) includes our data compression (rejecting modes with singular values less than zero). However, that highlights a material difference between the two methods because we use the data assimilation system to determine how much weight to give observations with singular values between zero and one while Migliorini et al. (2008) rejects those observations.

6. ***As discussed above, the correct form of the eigenvector decomposition of U_0^T E_m U_0 (equivalent to an SVD given the matrix is symmetrical, as noted by the authors) is nPhi nSigma nPhi^T not nPhi^T nSigma nPsi.***

Corrected. We only found one occurrence of this error (in the line immediately preceding Eq. 5). If there are others, please advise.

7. ***I agree with the authors' derivation. Their last expression can (should?) be simplified as (I – A) C_a A^T as noted in my previous comment.***

We have revised the text accordingly.

8. ***Revisions of paper to address Comments 2, 3, 4, and 5 and our response to this comments.***

We propose to revise two sections of the paper to address Reviewer 1's comments and our responses.

In Section 2 on page 6 at the line following Eq. (5) beginning with “This approach compresses . . .” and ending with “. . . the truncated identity matrix” we propose to replace that text with the following:

“Our approach compresses Eq. (3) so that the dimension of the “compact phase space retrieval” (CPSR) profile on the left side of Eq. (5) is identical to the number of independent functions linear functions of the atmospheric profile to which the instrument is sensitive. That method is different from that of Migliorini et al. (2008) because it compresses the quasi-optimal retrieval observations based on a linear independence analysis and relies on the assimilation system to decide how much weight to give the observations. Migliorini et al. (2008)’s approach reduces the number of observations based on an uncertainty analysis independent of the assimilation system. Our approach identifies all linearly independent information contained in the QOR profile (through projection of the QOR profile onto the left non-zero singular vectors of the averaging kernel). Migliorini et al. (2008)’s approach may; (i) discard some linearly independent information because the left non-zero singular vectors of the observation error covariance are not necessarily a basis for the space of QORs; and (ii) discard some linearly independent information through their uncertainty analysis. Finally, our approach relies on two transforms: (i) a compression transform (based on the left non-zero singular vectors of the averaging kernel, and (ii) a diagonalization transform (based on the left non-zero singular vectors of the compressed observation error covariance). Migliorini et al. (2008)’s approach uses two diagonalization transforms – one based on the observation error covariance and a second based on the transformed forecast error covariance in observation space. Our diagonalization transform is analogous to their first diagonalization transform except we apply it to the compressed observation error covariance, and they apply it to the “untransformed” observation error covariance. As in Migliorini et al. (2008), the final form of our observation error covariance is the truncated identity matrix.”

In Section 8 on page 21 at the last line of the page replace “. . . pression transform.” with the following:

“pression transform. Nevertheless, our CPSR approach is different from that of Migliorini et al. (2008): (i) we perform two transforms – a compression transform and a diagonaliztion transform, they perform two diagonalization transforms; (ii) we identify and assimilate all linearly independent information observed by the instrument, they may discard linearly independent information – some because their transform vectors are not necessarily a basis for the space of QORs and some because their uncertainty analysis discards some information that lies in the range of their transformed averaging kernel; (iii) our diagonalization transform is analogous to their first diagonalization transform except we diagonalize the compressed observation error covariance and they diagonalize the “untransformed” observation error covariance; and (iv) we rely on the assimilation system to decide how much weight to give the transformed observations and require no information from the forecast ensemble, and they use the forecast ensemble to decide which observations to discard.”

The next sentence, beginning with “MOP CPSR maps in Fig. 5” should begin a new paragraph.