Response to Anonymous Referee 1

- **P 4 line 22: "better representing" I don't know what is meant here, please restate.**

  Changed the sentence for more clarity from

  "Such simple methods have the advantage of better representing the instrument measurement, and enabling more feedback on instrument performance."

  to

  "Such simple methods have the advantage of enabling more feedback on instrument performance by virtue of forcing the retrieval to derive certain information strictly from the measurement even when non-optimal."

- **P 7: forward model definition. I find this section confusing. The measurement vector y is described as the deviation in the absorption from that corresponding to $\mathbf{x}_u$, but that is not a measurable quantity and the statement is contradicted by Eq. (4). Near the bottom of page 8 it is claimed a valid choice for the reference profile is $\mathbf{x}_u = \mathbf{0}$, so that y is a deviation from zero. I believe this is correct in the end, and exploits the assumed linearity of the problem, but it is still not entirely clear to me, and I think the concepts should be better explained.**

  We agree with the reviewer. We have modified Eqn. 4 to include the noise term **epsilon**

  The point of the equations at the bottom of page 8 were to use the offset and scale choices that we have at our disposal in setting how $\mathbf{x}$ relates to the CO_2 profile (in units of ppm). Our specific choice was made based on making the SVD equations least complicated. In the revised version, we will include an additional equation to explicitly show that relationship and include a sample calculation for clarity.

  In the examples we show, we set the uninformative prior to be a 400 ppm uniform column. With $\mathbf{x_u}$ being zero, $\hat{\mathbf{x}}$ having all elements 0 corresponds to a uniform column of 400 ppm. An element of $\hat{\mathbf{x}}$ having a value of 0.02 corresponds to that layer in the atmosphere having a mixing ratio of $(1 + 0.02) \times 400 = 408$ ppm. Similarly, an element of $\hat{\mathbf{x}}$ having a value of -0.02 corresponds to that layer in the atmosphere having a mixing ratio of 392 ppm.

The element in $\mathbf{x_u}$ corresponding to the surface reflectance or signal level (x_0) also has degrees of freedom for the offset and scale. Thus, one can set $\mathbf{x_u}$ to be zero with no loss in generality.

We have revised the text at the bottom of page 8 to

"In the above equations, we have carefully exercised our choice in linearly mapping the physical world to $\mathbf{x}$ by setting

$$\mathbf{x}_u = 0$$

for simplicity, and scaling $\mathbf{x}$ such that $x_i = -1$ corresponds to the GHG concentration of the i^{th} layer in the atmosphere being zero. As per Eqn. 3, $\mathbf{F}(\mathbf{x}_u)$ is a constant, which can also be set to zero with no loss in generality. These sorts of transformations are fairly standard in the literature and make the equations less complicated. "

We have also expanded our description of the numerical simulation methodology pertaining to the SVD method:

"For the SVD approach, we set the uninformative prior $\mathbf{x}_u$ to be a uniform 400 ppm CO_2 profile and anchor our definition of $\mathbf{x}$ to it. From this, $x = -0.02, 0, 0.02$ would correspond to mixing ratios of 392, 400 and 408 ppm respectively. "

- **P 21 line 7: "to the create"**

  We have corrected this to "to create"

- **P 28 line 11: "bias-free estimate" This is not true in general, as the authors have themselves noted on P 4 line 24. It should be qualified or its applicability defined**

  We have changed the sentence from

  "...confirms the notion that the retrieval of a column mean using least-squares line fitting of an absorption spectrum yields a bias-free estimate of the $\mathrm{X_{GHG}}$, regardless of the shape of the profile used in the prior (which turns out to be uninformative)"

  to

  "...confirms the notion that the retrieval of a column mean using least-squares line fitting of an absorption spectrum yields an estimate of the $\mathrm{X_{GHG}}$ without incurring bias from the regularization or retrieval, regardless of the shape of the profile used in the prior (which turns out to be uninformative) "

Response to Anonymous Referee 2

- **Partitioning the profile into preset shapes is discussed in Tukiainen et al. (2016) for TCCON CH4, although the retrieved shapes were based on prior covariance. The error analysis in this paper should be compared to Tukiainen (2016).**

  Although we agree with the referee that a comparison of the error analysis with that of Tukiainen et al. (2016) would be useful, we feel it would be beyond the scope of this work. While there are some similarities between the methods, namely that principal components with truncation are used to solve the problem, there are two key differences. The first, as the referee has noted, is that Tukiainen used the prior covariance to determine the terms in the dimension reduction. The second key difference is that we report results in the principal component basis ($\mathbf{z}$) rather than full model space ($\mathbf{x}$) for the purpose of either GHG flux modeling and retrieval validation.

  The SVD method as described in our work has the key advantage of being able to retrieve scientifically useful quantities with any bias from the regularization process even in the absence of prior information about the GHG profile. Projecting such results onto the full model space will require information from the prior and thus introduce bias, since the prior is meant to be uninformative.

  Nevertheless, it will be useful to make a one-to-one retrieval and error analysis comparison on a TCCON like system between our use of the SVD method and that of Tukiainen et al (2016). While that is not within the scope of this work, we will definitely consider it for future work.

- **One of the main ways this approach is validated is by comparisons to optimal estimation, however the optimal estimation retrievals do not look comparable to profile retrievals from OCO-2 available in the L2 standard products. For example Figure 6 shows 5 oscillations in the retrieval on the order of 50 ppm. OCO-2 retrieved profiles do not show these types of oscillations. It appears from p. 24 line 1-2 that the constraint used in OE is diagonal. The constraint used should match ODell, 2012 (Figure 2) which has strong off-diagonal correlations. Comparing the SVD retrievals to state of the art OE retrievals will be useful**

  We agree that a comparison of the SVD retrievals to the state of the art OE retrievals will be useful. However, there are several complications

that go into the choice of a Bayesian prior such as that used in OCO-2, such as local meteorology, vertical mixing and confidence in global GHG models at the location in question. Our intent behind this work was to showcase the SVD method and compare and contrast it with the OE method using a simplified system. Hence, we decided to use a conservative, 200 mB $1/e^2$ vertical correlation distance in the CO_2 mixing ratio in the atmosphere for off-diagonal terms (Page 22, line 9).

We do plan future work to make a comparison between the state-of-the-art OCO-2 retrievals and one based on the SVD method.

In the paper, we have expanded on the description of the Bayesian prior chosen: "For the OE approach, a proper choice of a Bayesian prior would factor in local meteorology, vertical mixing and confidence in global GHG models at the location in question. However, for the purpose of illustration of the workings of the OE method, we have kept the Bayesian prior mean and covariance simple. The Bayesian prior mean and variance (diagonal terms on the covariance matrix) are chosen on a case-by-case basis. For the prior covariance (off diagonal terms in the covariance matrix), we assume a 200 mB $1/e^2$ vertical correlation distance in the CO_2 concentration in the atmosphere. "

- **The second issue in this paper are the claims in the abstract that SVD results in unbiased results and is therefore better than OE.**

  It was not our objective to show that SVD is "better than OE". In Section 3.3, we showed that SVD is equivalent to an OE estimator with an uninformative prior and a Moore-Penrose pseudoinverse. That is, SVD can be considered as a subclass within the OE framework. Also, as seen in Figure 12, we specifically advocate the OE method when good quality prior information is available since it gives the best estimate.

  In the paper, we drew a distinction between the types of priors used within OE (informative) and within SVD (uninformative), and we derived some interesting properties of the two choices of priors. They each have their own strength, which we summarize in a new paragraph in the Summary section:

  "Intuitively, OE derives an estimate of the state using both the measurement and prior knowledge, while SVD only uses just the measurement to inform its estimate. When the prior information is correct, there is no doubt that OE will have lower posterior uncertainty since

OE can leverage an extra source of information to more efficiently derive its estimate. However, this efficiency comes at a potential cost when the prior is *incorrect*. For instance, we showed that when OE uses an incorrect prior mean, then the estimate is guaranteed to be biased. Estimates from the SVD method in the principal component basis, on the other hand, are insensitive to incorrect information coming from the prior. The choice between SVD and OE then mostly comes down to how well one understands the prior distribution of the state of interest."

- **While it is true that the basis functions may not need to be constrained if truncated at whole degrees of freedom, and there may be no biases in the mapped space, the translation of the basis functions into a profile can result in biases and these should be quantified.**

  The translation of the basis functions into a profile can indeed result in biases. We did show analytic expression for the bias of OE and SVD *in the original profile space* in the paper. They are in equation (30) and (33). For your convenience, we include them in this response. Suppose that OE and SVD both uses a wrong prior mean $\mathbf{x}_b$, which is different from the true prior mean $\mathbf{x}_a$, then the expected bias for OE is

$$Bias_{OE} = (\mathbf{I} - (\mathbf{S}_a^{-1} + \mathbf{K}\mathbf{S}_\epsilon^{-1}\mathbf{K}')^{-1}\mathbf{K}\mathbf{S}_\epsilon^{-1}\mathbf{K}')(\mathbf{x}_b - \mathbf{x}_a). \tag{1}$$

  And the SVD bias is

$$Bias_{SVD} = (\mathbf{I} - (\mathbf{K}\mathbf{S}_\epsilon^{-1}\mathbf{K}')^{+}\mathbf{K}\mathbf{S}_\epsilon^{-1}\mathbf{K}')(\mathbf{x}_b - \mathbf{x}_a). \tag{2}$$

  Note that when OE uses the correct prior mean ($\mathbf{x}_b = \mathbf{x}_a$), then $Bias_{OE} = 0$. Also, when $(\mathbf{K}\mathbf{S}_\epsilon^{-1}\mathbf{K}')$ is invertible, then $Bias_{SVD} = 0$ regardless of the choice of $\mathbf{x}_b$.

  Detailed error analysis with simulations was done in the $\mathbf{z}$-basis because the full GHG profile ($\mathbf{x}$-basis) is often not needed for use in GHG flux modeling. The retrieved parameters obtained from the $\mathbf{z}$ basis retrievals can often be directly mapped to a GHG column mean and other higher order components. It should be noted that this is the basis of line-fitting methods even if they don't explicitly use the SVD method, since they derive X_{GHG} strictly from the measurement, whose

information is contained within the SVD basis. X_{GHG} is ingested or assimilated into GHG flux models today. Higher order components like the vertical dipole moment can also be similarly ingested based on their information content as has been described in Joiner and Da Silva, *"Efficient methods to assimilate remotely sensed data based on information content"* (1998). A reference to Joiner and Da Silva (1998) has been added. A sentence has been added to the end of the last paragraph in section 2.1 "Joiner and Da Silva (1998) describe a method that can ingest such components into an assimilation model based on their information content. "

- **The biases introduced by this approach should estimated by calculating the linear estimate for different true states, e.g. Tukiainen (2006) Fig. 3 shows the difference between AirCore and smoothed AirCore for methane and a similar SVD approach. The column difference between AirCore and smoothed AirCore (or some other set of trues) would give the bias and error resulting from the SVD mapping using Eq. 34 from this paper.**

  The intent of the SVD approach described is for functional retrievals that can provide inputs into GHG flux models either in the absence of prior GHG profile information or when such information is of unknown quality or potentially biased. As stated above, these do not require a retrieval of a full vertical profile. Nevertheless, we have shown the expression for calculating the error in the GHG profile. Please see Eqn (1) and (2) from the response above.

- **Section 3.3 is also hard to follow.**

  We rewrote section 3.3 slightly to indicate that the OE under certain assumptions is equivalent to SVD . Specifically, we rewrote the equation block in (22) to indicate that the retrieved value $\mathbf{x}_{uOE}$ arising from OE with an uninformative prior and pseudo-inverse is *identical* to the retrieved value $\mathbf{x}_{SVD}$ arising from SVD.

- **The authors should clarify how many basis functions are selected. If there are 1.6 degrees of freedom, are 2 basis functions used? If 2, wont the retrieval need some constraint? If 1, wont that throw away information? Kulawik et al. (2017) used 1.6 degrees of freedom from GOSAT to get 2 parameters each with about 0.8 degrees of freedom (so some a priori component in the retrieved values). Would this approach be able**

to get any vertical information with 1.6 degrees of freedom?

The calculation of degrees of freedom depends on the state of prior knowledge of the true state. In the SVD framework as we have used it, with prior information being absent, the retrieval process works the same whether the prior GHG profile is well known or poorly known. The key metric is the posterior uncertainty of the retrieved quantities. One can compare the posterior uncertainty against that of the prior knowledge of the system.

With retrievals in the principal component basis, the retrieved parameters are orthogonal to each other. The practical implication of this is that the mean and variance of a retrieved parameter is independent of the inclusion or exclusion of higher order parameters. This is illustrated in figures 10 and 11.

If a higher order parameter is retrieved despite insufficient information, that parameter will have a posterior uncertainty that is too large to be useful. If 2 degrees of freedom are retrieved when the Bayesian framework projects 1.6 degrees of freedom, the second order component will have an uncertainty that is slightly too large to be useful on its own. As mentioned previously, it will not affect the retrieval otherwise.

- **I do not follow Figure 4. What are the units on the y-axis? Are the authors aware that OCO-2 has better precision and more degrees of freedom than GOSAT? This figure suggests the opposite.**

  We are aware that OCO-2 has better prevision and more degrees of freedom than GOSAT. The purpose of Figure 4 was to illustrate the effect of spectral resolution, particularly on the higher order components. The overall precision depends on several factors including the light gathering capacity, detector sensitivity, integration time, etc.

  We have replaced the words GOSAT and OCO-2 with regions indicating spectral resolutions of satellite instruments. Here is the new Figure 4 with caption

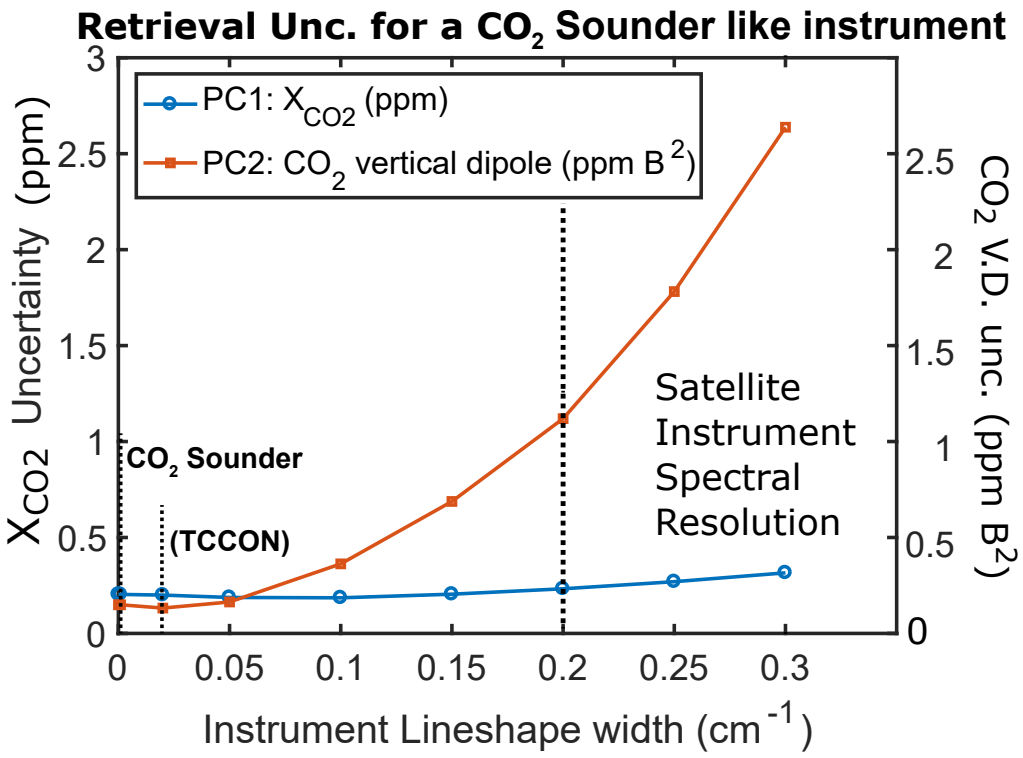

Figure 1: (Figure 4 in manuscript) Retrieval uncertainty versus instrument spectral linewidth for the first two CO_2 principal components (PCs) : While the column X_{CO2} is largely unaffected by the spectral resolution, the precision of the CO_2 vertical dipole moment degrades strongly with poorer resolution. We assume a CO_2 instrument model, but with some instrument line broadening. The x-axis denotes the full-width at half maximum of the triangular instrument line shape used to broaden the CO_2 absorption. We assume photon shot noise with a SNR of 1000 for points with no CO_2 absorption. The spectral resolutions of TCCON, satellite GHG sensing spectrometers and the CO_2 Sounder instrument are indicated, though the calculations done in this work apply only to the CO_2 Sounder instrument.

Other changes

- Caption on figure 12. We will incur a bias from the SVD regularization if the column mean is not a principal component
- We have replaced the use of the word "bias-free" in relation to the retrieved SVD components with a more accurate description, *i.e* that the components do not incur bias from the regularization process or

the use of an uninformative prior.