Review of the discussion paper "A singular Value Decomposition..." by A.K. Ramanathan et al.

As already stated by Reviewers 1 and 2, this paper tackles an important issue and I think that its topic fits perfevtly in AMT. In the following I will discuss some of the issues raised by the preceding reviews and will add a few comments of my own.

Rev. 1:
Comment:
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

Reply:
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 XGHG, 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 XGHG 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)"

My view:
As stated below in more detail, this is not generally true but only if the signal measured is independent of any other altitude-dependent atmospheric state variable. Care should be taken to avoid any generalizing statement on bias-freeness but to limit the statements to particular results of the investigations made.

Rev. 2:
Comment:
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).

Reply:
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 (z) rather than full model space (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.

My view:
There are many retrieval methods which aim at minimizing the bias of the retrieval by avoiding informative prior and which keep the result stable basically by reducing the effective dimension of the result vector. Tukiainen (2016) is one of these, von Clarmann et al. (AMT 8, 2749-2757, 2015) would be another one, and I am sure there are a lot more. Comparison with each of these methods would be interesting but a line must be drawn somewhere and the decision which of these comparisons should be included in the paper should be left to the authors. The authors should not be forced to discuss any specific one of these papers unless a compelling reason is given that paper x is more important than paper y.

Comment:
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.

Reply:
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/e2 vertical correlation distance in the CO2 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/e2 vertical correlation distance in the CO2 concentration in the atmosphere."

My view:
Again I agree with the authors. The purpose of the related section is validation, not competition. A common practice of validation is comparison with something well understood and easily traceable. The choice of a simplified OE approach appears to be rational and justifiable to me.

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

Reply:
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.

My view:
While I agree with the argument of the authors in its heart, I have some reservations with the use of the term "OE" in the context of ad-hoc priors or uninformative priors. See my own comments below.

Reply (cont'd):
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."

My view:
Bayesian and non-Bayesian methods just answer two different questions. Bayesian methods (with realistic prior) tell us what the most probable state of the atmosphere has been, while non-Bayesian methods tell us what the most plausible interpretation of the measurement is. None of these is superior; different questions demand different answers. But a Bayesian formalism with uniformative prior or ad hoc prior, I would say, are a variant of constrained maximum likelihood retrievals in disguise. I admit that Bayes himself endorsed uninformative prior, and also Gauss did it, but this has been heavily criticized by Pearson and Fisher, and we cannot be sure what Bayes himself thought about this issue because he did not publish his work while still alive.

Comment:
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.

Reply:
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 xb, which is different from the true prior mean xa, then the expected bias for OE is BiasOE = ... And the SVD bias is BiasSV D = ... Note that when OE uses the correct prior mean (xb = xa), then BiasOE = 0. Also, when (KSKt-1 is invertible, then BiasSV D = 0 regardless of the choice of xb. Detailed error analysis with simulations was done in the z-basis because the full GHG profile (x-basis) is often not needed for use in GHG flux modeling. The retrieved parameters obtained from the 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 XGHG strictly from the measurement, whose information is contained within the SVD basis. XGHG 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 DaSilva, "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."

My view:
I still think that, in the general case, truncated SVD can cause a bias, even in

the z-space. See my own comments below.

Comment:
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.

Reply:
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.

My view:
Here the authors miss the point the different states result in different K-matrices. I think this is why the reviewer wants to see linear estimates for different true states. On the other hand, I have no idea how state-dependent the K matrix is in the given application. Possibly an adequate caveat could save the paper without much additional investigation.

Comment:
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, won't the retrieval need some constraint? If 1, won't 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?

Reply:
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.

My view:
A naive question: Why not considering only a fraction of the critical component, e.g., when reconstructing the profile in the x-space, on might consider the first singular vector in full and the second only with reduced weight? But such a modification of the method - if possible at all - would be beyond the scope of the paper. Given the application of the data intended and discussed in the paper, the authors' argument is compelling.

Comment:
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.

Reply: 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 CO2 principal components (PCs) : While the column XCO2 is largely unaffected by the spectral resolution, the precision of the CO2 vertical dipole moment degrades strongly with poorer resolution. We assume a CO2 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 CO2 absorption. We assume photon shot noise with a SNR of 1000 for points with no CO2 absorption. The spectral resolutions of TCCON, satellite GHG sensing spectrometers and the CO2 Sounder instrument are indicated, though the calculations done in this work apply only to the CO2 Sounder instrument.

My view:
The purpose of this part of the paper is a methodical study, not a comparison between two existing instruments. I consider the action taken by the authors in reply to the revieM as adequate.

Comments:
I look forward to reviewing this paper when the authors improve the OE results and more carefully characterize the bias and errors compared to the current OE method.

My view:
The authors and the reviewer seem not to agree what the scope of the paper shall be. If the paper is understood as a methodical study and not as an evaluation of existing instrument concepts, the criticism by the reviewer appears a bit too harsh to me.

Other changes by the authors (I do not comment on all of them):

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 7the use of an uninformative prior.

My view:
This does not fully solve the problem; this is because K can depend on x. A different x_u can thus still lead to a different result. See also my own comment below. What is needed here is either an assessment that the dependence of K on x is weak enough to be neglected or a caveat that this type of effects can exist but has not been assessed.

My own comments (pagination refers to amt-2018-14-manuscript-version3.pdf):

p2 l16 Hansen (1990) -¿ (Hansen, 1990)

p2 l17 The statement that the retrieved principal components are unbiased or bias-free may be true in the given context but not in general. Assume a thermal emission instrument where the signal depends on the temperature of the emitting layer. The SVD method will change the profile shape by removing fine structure. With this a certain amount of gas may be shifted into another altitude where the temperature is different. Thus, the same amount of gas may generate a different amount of radiance, and in turn, the retrieved total amount of gas depends on the assumed profile shape. This counter-example may sound contrived but at least it disproves the general validity of the statement made. This mechanism may become effective only within an iterative context. But even without considering an iterative process, K depends on x_u, even if the uninformative prior is realized by setting S_a^{-1} zero.

p2 l30: The term "uninformative prior" plays an essential role in this paper, thus it needs to be clearly defined when first used.

p4 l12 Not clear why least squares fit is needed if the retrieval problem is fully determined. A direct solution by matrix inversion would be possible. Least squares fit is only needed if the problem is OVER-determined. It is clear what the authors mean but the wording is a bit sloppy here and may direct the reader into a wrong direction.

p4 21-27 These statements are certainly correct but usually this problem is solved by distinguishing between "variables" and "parameters" of the retrieval problem; those input values of the forward model which are kept constant during the retrieval can be called parameters and those which are part of the x-vector are the variables. I think this would comply with traditional language in the case of a function F with multi-dimensional input. The term 'a priori' then could be reserved for the latter ones. This terminology would make this paragraph obsolete but would be in conflict with the terminology in the remainder of the paper (e.g. Fig 1 where what I call 'variables' is called 'parameters'. Probably it is the best to leave the terminology as it is, because it is, at least, self-consistent.

p5 l13 (I know that I am exaggeratedly fussy with such issues!) I do not quite agree that one really can gradually move back and forth between Bayesian and non-Bayesian methods. The reason is this: The solution of a Bayesian retrieval can be interpreted in terms of probability. The solution maximizes the a posteriori probability. A method which does not use the prior information will not render a solution which can be construed as the maximum of a probability distribution but should be understood in the sense of likelihood (Fisher, 1922). The move from 'conceivable as probability' to 'not conceivable as probability' is discontinuous, even if formally and result-wise the move is continuous. I think the problem can easily be remedied by writing "...from fully-Bayesian-like formalism to ...". Then it is clear that there is no transition between the concepts behind the formulas and no claim is then made that the Bayesian formalism really represents the concept of maximization of the a posteriori probability.

p5 l11: (I am still fussy...) Is there really a "choice of the prior covariance matrix". I know that this terminology is often used as internal slang of the retrieval community; but is, in a Bayesian sense, the a priori covariance matrix really something one can 'choose'? I suggest a slightly modified language where, whenever an ad hoc choice of the a priori is used, the term 'a priori ASSUMPTION' is used, and a ... matrix IN PLACE of the a priori covariance matrix.

Eqs 16-22 and related text: What you actually show (because you set ${}_{S}a^{-1}$ to zero) that the SVD retrieval is equivalent with a simple Gaussian weighted least squares retrieval admittedly, Gauss also used a non-informative prior to give this approach a probabilistic interpretation. This comes down to a maximum likelihood retrieval (Fisher, 1922). I find it misleading to claim to have shown the equivalence with an OE retrieval, because the main characteristic of the latter as understood today is that it does use INFORMATIVE prior.

Sect 3.5. See my comment above: You formally prove that the SVD method is bias-free in a sense that the SVD concept does not introduce a bias and that the prior remains ineffective. This seems to be in conflict with the bias-causing mechanism I propose above. The reason for this conflict is this. The formal proof uses linear algebra. My proposed mechanism is nonlinear because it considers effects which are caused by the profile-dependence of K. The SVD

tyoically removes fine structure of the profile. Imagine the the atmosphere is particularly hot where a peak in the profile is located. Let SVD remove this peak, in a way that the total column is unchanged. With the same amount of molecules, the forward model now produces less radiance, and the retrieval will put in additional molecules to compensate for this. In other settings than thermal emission, e.g. reflected solar radiance, the effect may be less pronounced but absorption cross-sections are also temperature dependent; and there may be further mechanisms which might cause a profile-dependence of K. The truncated SVD approach changes the profile (in the x-space to which z has to be transformed back if F is to be evaluated in the second iteration) this approach may cause a bias. And as said above, even in a non-iterative context, K depends on x_u, even if Sa^{-1} is set to zero. Thus, via K, the prior is not as uninformative as it may appear. It was very audacious to claim that the method is always bias-free, and even the wording in the revised version still appears too strong to me. I suggest to add a caveat to the text, like "within linear theory, i.e. without consideration of the profile-dependence of K" or something similar.

Sect 4.4.: The concepts of resolution and sampling are often confused. I thus appreciate that here it is between both these concepts. However, I have two suggestions to make this paragraph even clearer.
1. p19 l20: Here a 'smaller' resolution is mentioned. Language is (in general, not only here) confusing because if a small number is associated with the resolution, the resolution is good and if a large number is mentioned, the resolution is poor. When a 'small' resolution is mentioned, it is not clear if the resolution is good (low number) or if the resolution is poor (high number, low resolving power); a terminology using good vs. poor resolution would less unambiguous.
2. p19 l22: I suggest to avoid to combine the terms 'resolution' and 'sampling', because these concepts are too often confused. Thus I suggest to replace 'high sampling resolution' with 'dense sampling'.

Sect. 6.4. The most relevant difference between the Bayesian method (with an a priori which correctly describes the background statistics) and other methods is that it does, contrary to all other methods, renders the most probable solution. Methods which do not, in one way or another, invoke the Bayes theorem don't do this.

By the way, SVD is not the only rationally founded way to reduce the effective dimension of the retrieval. The common goal is to make the retrieval stable without becoming explicitly dependent on prior information. Which of these methods is the most adequate depends on the intended application of the data, and I appreciate that the authors raise this issue. A particular disadvantage of OE (with informative prior) is that the averaging kernels depend on the atmospheric state and thus on time, which makes time series hard to understand and interprete. Similarly, the singular vectors of an SVD method depend on the atmospheric state. Von Clarmann et al. (2015) suggest a different method where the retrieval basis is time-independent (but this may be a bit off-topic

here).

SUMMARY:
In summary, it seems to me that all issues can be remedied by purely redactional actions without much further scientific inverstigation. Thus I recommend publication after minor revision.