The authors thank the reviewer for his comments.

Review #2:

(2.1) **Review: I have to confess that I am still puzzling what was the real intention of the authors in submitting this long and, to some extent, verbose report for publication to AMT.**

Reply: The intention of this paper is to summarize retrieval approaches actually in use in satellite remote sensing, to systematize them in a common framework and notation, to discuss the implications of related choices on error propagation and to infer related recommendations on unified error reporting.

Planned Action: We will state this intention clearer in the introduction

(2.2) **Review: Although I appreciate the effort in contributing to simplify the exchange of L2 data and explain their error characteristics, in its present version the paper seems just an occasion for the many authors to recount and selfreference what they did in the area of inverse/retrieval algorithms for the sounding of atmospheric parameters.**

Reply: The purpose of the article is to cover all the satellites for remote sensing of atmospheric compositions over the past 20 years from the all frequency ranges, microwave, infrared, NIR and UV/VIS, as a review paper. This implies numerous references, and since the list of authors includes many scientists from many different groups working in this field, it appears quite natural to us that self-referencing is inavoidable.

Planned Action: We will add more references from scientists who are not involved in this paper in order to make the article more balanced.

(2.3) **Review: The title seems to open to a wide tutorial, however at the end of the abstract they say the goal of the paper is just to provide a list of recommendations which shall help to unify retrieval error reporting.**

Reply: We are afraid that this is misreading; we do **not** say that the goal of the paper is just to provide a list of recommendations. We do mention that we provide some recommendations, but the rest of the abstract summarizes the problem areas tackled in this overview paper. Only in the section on conditions of adequacy we indeed mention the "ultimate goal of presenting a list of recommendations". The attribute "ultimate" makes clear that this is by no means the only goal.

Planned Action: We will address this more clearly in the abstract, introduction, and top of the recommendations.

(2.4) **Review: In section 3, it seems that the authors want to redefine terminology about errors. Do we have to call the root mean square error, simply uncertainty? And the variance, precision? Or whatsoever? Do we have to stick to new definitions issued by JCGM and BIPM? Is it a problem of terminology or contents? Or simply, do authors want to set up a sort of protocol for exchanging L2 products?**

Reply: We want to avoid quibbling about words. The reviewer is free to call the quantities mentioned as they like, as long as the terms are clearly defined somewhere. However, the terminology we use is applicable also to single measurements, while we have problems to assign a meaning to the terms 'bias' or the 'root mean square error' in the case of a single measurement.

Planned Action: none

(2.5) **Review: By the way, in the end, I count 6 CoAs and 18 (with subpoints) recommendations, for a total of 24 and more. To me, more than 3 recommendations are effective as no recommendations at all. In effect, 24 recommendations are normally much more than the degrees of freedom or pieces of information conveyed by common retrievals.**

Reply: We do not understand how it is logically justified to calculate the sum of the CoAs and the recommendations. We thought that sums can only be calculated of items of the same category. We also do not understand what the logical link between the number of recommendations and the degrees of freedom of a retrieval is. We do not see how are these quantities connected. To us, these quantities seem imcommensurable.
We would have preferred less recommendations but condensing them makes them less specific, and finally we would end up with some vague truisms which would not be helpful at all.

Planned Action: none

(2.6) **Review: Looking deep inside the paper, I can see interesting aspects about trying to define a common paradigm to interpret data coming from a large variety of satellite data processors. However, this objective is somewhat lost among unnecessary details of retrieval schemes, methodological issues [...]**

Reply: The retrieval schemes and methodical issues belong to the core content of the review paper. Without understanding the underlying simplifying assumptions of a retrieval scheme, it is difficult, perhaps even impossible, to provide a reliable error estimate.

Planned Action: The introduction will be rewritten to make the purpose of the paper clearer.

(2.7) **Review: [...] and what I could call a silent but insistent criticism to Optimal Estimation.**

Reply: We neither endorse nor dispraise any particular method but we describe the methods which are currently in use, or whose data products are currently in use. For each method we discuss the underlying assumptions.

Planned Action: We will make an explicit statement that the superiority of either maximum-likelihood based or optimal-estimation based retrievals cannot be decided on scientific grounds but is a purely philosophical question.

(2.8) **Review: Furthermore, I think that the format of the present study is much more adequate for a report.**

Reply: Reports typically report technical information related to one instrument, processor, etc. We present, in a unified notation, an overview of all methods we are aware of. Thus we think that this paper serves well as an overview paper for the TUNER special issue, because it provides a framework the other papers of the special issue can refer to.

Planned Action: none

(2.9) **Review: Concerning retrieval error reporting, the canonical Theory of Statistics has been teaching us (e.g. Kendall and Stuart Vol I, II, III, The Advanced Theory of Statistics, Fourth Edition, 1979) for so many years that the performance of a given statistic or estimator, say $\hat{x}$, is measured in terms of its mean square error or deviation from the true value, which can be decomposed in variance and bias, namely**

$$E[(\hat{x}-x)^2] = E[(\hat{x}-\bar{x})^2] + E[(x-\bar{(}x))^2]$$

For the assessment of the root mean square error and its reporting, the consolidated usage is today to share and/or distribute.

1. **Estimated state (of course) and related retrieval covariance matrix**
2. **Background (state and covariance)**
3. **Averaging Kernels**

Based on the above items, the performance of any estimator (bias and variance) can be unambiguously quantified. From what I can see, in the end, the above three ingredients are what authors agree with to be the basic items to share. In this respect, a potential list of recommendations, included that of authors, could be made and

explained in onetwo pages.

Reply: First a side remark: The fact that canonical theory of statistics relates the performance of a statistic estimator to the true value strengthens our position against GUM. We do agree that the errors of an ensemble of retrievals can be decomposed into the mean square error and the bias, and we use this concept ourselves in order to validate the error estimates. We concede that our list of recommendations is three pages, but it covers issues not mentioned by the reviewer (correlations in other domains; data traffic, and others).

Planned Action: none

(2.10) **Review: I have also to say, that authors' recommendation list itself is largely independent of the bulk of the present paper.**

Reply: We admit that not all parts of the paper are needed to derive the recommendations, but the information contained in the bulk of the paper is needed to provide the quantities requested by the recommendations.

Planned Action: The relation between the recommendations and the rest of the paper will be made clearer.

(2.11) **Review: General Remarks**
The paper is lacking a correct definition and assessment of bias. Authors seem to identify the random component of the root mean square error as the error or uncertainty of a given retrieval system. What about the bias? What's the strategy they want to set up to estimate it and eventually share with end users?

Reply: It is not true that we identify the random component of the root mean square error as the error or uncertainty of a given retrieval system. We conceive the error or uncertainty as a quantity which is composed of a random part (corresponding to what the reviewer calls root mean square error) and a systematic part (corresponding to what the reviewer calls bias). We state explicitly that the systematic error estimates can be tested using the bias between collocated measurements of independent measurement systems, and that the random part can be tested using the standard deviation of the difference between collocated data from different measurement systems. The bias is commonly defined as a mean difference between the measured value and the true value unless explicitly specified differently. In our paper, when we use the term 'bias' with any other meaning than the mean difference between the measured and the true value, we state explicitly what the mean difference refers to.

Planned Action: We will add a paragraph on the bias.

(2.12) **Review: I have found a bit confusing the question about gridin-**

dependent retrievals, which for me is a nonsense, since normally one works with a discretized state vector. Apart from forward model (FM), the bias depends on the given constraints, which are normally griddependent, in the sense that their definition and use is contingent to the way the state vector has been discretized. In effect, for a regularized estimator the bias depends solely on the constraints (again apart from FM biases).[...]

Reply: Another contribution to the bias can be calibration issues. The role of the constraint is discussed in Section 6.

Planned Action: We will mention that the choice of a prior which is not the expectation of an ensemble the actual measurement is taken from will cause a bias.

(2.13) **Review: This basic aspect has been largely overlooked in the paper, and in fact their recommendations are not consistent with a correct sharing of the root mean square error.**

Reply: We recommed that the averaging kernels and priors used shall be communicated to the users. The users can then evaluate the smoothing error on the final grid they use, after evaluating the additional averaging kernel component entailed by the interpolation. Sharing the total error will cause inadequate error estimates after resampling and respective generalized Gaussian error propagation.

Planned Action: As said above, the discussion of biases will be expanded.

(2.14) **Review: On the same line, their CoA2 is inconsistent with the idea of root mean square error.**

Reply: The intention is to avoid that data users interpolate the smoothing error on a finer grid. Instead they should be provided with all information they need to directly evaluate it on the grid of their choice. Any possible inconsistence with the root mean square error comes only from conceiving the retrieved state as a smoothed estimate of the truth, a conception we do not explicitly endorse. Conceiving the retrieved state as an estimate of the smoothed truth removes this inconsistency.

Planned Action: none

(2.15) **Review: Furthermore, I am not sure if it can be implemented, in practice.**

Reply: For noise alone, CoA2 can be implemented. It is only the combined noise and smoothing error which causes the problem.

Planned Action: none

(2.16) **Review: To streamline my personal thinking, let's suppose W is a suitable interpolation/extrapolation operator, which transforms a given estimator $\hat{x}_{n1}$, defined on a grid with n_1 layers, into a new one, say $\hat{x}_{n2}$, defined on a grid withn_2 layers, we have**

$$\hat{x}_{n2} = W\hat{x}_{n1},$$

with W a matrix of size $n_1 \times n_1$. CoA2 requires that, using authors' language,

$$WS_{x,noise,n1}W^T = S_{x,noise,n2},$$

where, $S_{x,noise,XX}$ is the error covariance directly retrieved on the grid with XX layers. However, I cannot see how the above condition can be met for any choice of W and $n_1 \leq n_2$ or W and $n_1 \geq n_2$. Atmospheric state vectors are not bandlimited signals, therefore a mere extrapolation/interpolation of a given retrieval from a coarser to a finer grid will not show finer structures of the underlying state. Hence, the above condition would normally not be met.

Reply: We agree with everything above except that the additional error is not part of the noise but of the smoothing error, which, we suggest, should be evaluated newly on the finer grid. The example presented by the reviewer shows perfectly why we insist that noise and smoothing error should be reported separately. For noise alone, CoA2 is fulfilled by using generalized Gaussian error propagation. And again, conceiving the retrieval as an estimate of the smoothed truth removes this inconsistency.

Planned Action: none

(2.17) **Review: Condition CoA2 seems to have been set up just to criticize the concept of smoothing error, which is the way Rodgers considers for the bias. Since the bias of the individual, single, retrieval depends on the true value, which is normally not know, Rodgers considers the variability of the true value (variancecovariance) in order to have at least an estimates of the interval in which the bias is expected to range. However, the variability and/or stochastic behaviour of the state vector, which is correctly considered in OE, is overlooked by authors.**

Reply: We do not agree. We do not criticize the concept of the smoothing error in general (except for the ambiguity of the underlying interpretations of probability, which we criticize in a very careful and moderate wording). The central point of our criticism is the inclusion of the smoothing error in the total error, which will lead to inconsistent results after resampling of profiles.

Planned Action: none

(2.18) **Review: They say, "natural variability is not a genuine retrieval error". It seems to me that authors purposely mislead statistical error with mistake. Natural variability is what makes our weather to be forecastable, but not exactly predictable. This is why we need statistics to address natural variability.**
Taking into account the natural variability of the state vector, it is possible to perform an assessment of the estimator's bias, e.g., through the (unfortunately named) smoothing error, whose meaning has been, in fact, completely mislead by authors (see also later when dealing with the smoothing error).

Reply: To us, natural variability explains that the atmospheric state at one time and one place is different from the state at another place and another time. Due to this natural variability we cannot expect that two instruments measuring at different places and/or times will render the same result. Detected differences thus do not hint at any malfunction of one of the instruments or retrieval and thus are not genuine measurement errors. Still, these differences have to be considered in comparisons. The reviewer has torn this quotation out of a very different context in our paper. From the context of Section 6.6, where the quoted statement comes from, it should be very clear what we mean. We do not understand how the reviewer can, on the basis of this text, accuse us to "purposely mislead statistical error with mistake".

Planned Action: We will add "[...natural variability] in a sense that the atmospheric state at place s_1 and time t_1 differs from the one at s_2 and t_2." And we will give more weight in the text to the regularization bias.

(2.19) **Review: Finally, because of the many issues addressed in the paper, in the end it looks like a confusing revision of Rodgers 2000; a sort of poutpourri of about everything is known today on atmospheric inverse problems: Twomey, Tikhonov, Rodgers, LS, ML.**

Reply: Our intention is to cover all relevant (in the sense that data retrieved with these methods are still around) methods within a consistent framework and a common notation. This is a precondition for unified error reporting. While the book of Rodgers (2000) provides an excellent theoretical basis, we apply this theory (and other variants) to the real-world retrieval schemes and investigate which uncertainties are caused by the assumptions and approximations in place. We understand this as a systematic compilation rather than a potpourri. We first lay down the basic theory. Then we discuss how retrieval schemes used in the real world deviate from the idealized theory. Then we discuss all error sources and their relevance. We find that the content is clearly structured, and goes beyond the content of the available literature in that it treats also the relevant real-world problems.

Planned Action: The introduction of the paper will be rewritten to make the purpose of the paper more evident.

(2.20) **Review: Furthermore, the estimator described in Eq. (4) in the text is not rigorously derived from any basic principle of statistics, it is just copied from OE and rewritten by substituting S_a^{-1} with R**

Reply: From our introduction it should be clear that we do not only consider methods which have a probabilistic interpretation. T. von Clarmann and U. Grabowski, Atmos. Chem. Phys. 7:397-408 (2007), their appendix, have shown that there is even a probabilistic interpretation of Eq 4 with R defined as shown in Eq 5. We do not see what is wrong with putting a method in a more general context.

Planned Action: none

(2.21) **Review: Specific Comments**
Pag. 3. At best, CoA2 is only consistent with the variance component of the estimated error. What they want to do with the bias is not clear. Stand as is, I have doubt CoA2 is effective and can really work.

Reply: Resampling of profiles and associated error propagation works well for all error components (noise, instrumental calibration biases, forward model biases...) except those which depend on the sampling of the $\mathbf{S}_a$ matrix. Thus we insist that the latter should be evaluated on the final grid, using the respective sufficiently resolved covariance matrix.

Planned Action: none

(2.22) **Review: Page 4. Section 3.1 This is confusing. Please state exactly why uncertainty cannot be used or why it sounds ambiguous if referred to the root mean square error of an estimator.**

Reply: We do not say that 'uncertainty' shall not be used. We say that the claimed difference between 'uncertainty' and 'error' is controversial. And according to GUM (and with respect to this issue we agree with GUM), 'uncertainty' does NOT refer to the root mean square error of an estimator but includes also systematic effects.

Planned Action: none

(2.23) **Review: Page 6, Eq (3), I cannot see any point why the unconstrained Least Squares solution should be called "Maximum Likeli-**

hood". This is a misconception. The assumption of Normal pdf is what really qualify the estimator (3). The reason of using ML because it yields LS under normality is untenable; it is like saying that a meteorologist is using Einstein General Gravity (EGG) theory when forecasting the atmosphere with the Newton dynamical equations, because EGG retrieves Newton in the limit of low velocity.

Reply: The term 'maximum likelihood' in this context is used by Rodgers (2000) for a solution which is free of formal prior information. And this terminology is consistent with that of Fisher, who coined that term. If we search for a solution of which the probability that it reproduces the (noisy) measurement is largest, we get, by definition, the maximum likelihood solution. If we apply this principle to Gaussian noise, the maximum likelihood solution happens to be the least squares solution. We do not use the maximum likelihood solution because it yields LS under normality but we use least squares because ML plus normally distributed yields least squares. It is agreed - and even conceded by Fisher - that ML does not yield the solution of maximum posterior probability. But what is untenable about it? we do never claim that we consider only methods which have a probabilistic interpretation. And more generally speaking: We do not particularly endorse any of the methods we describe. In this paper, we just describe and characterize them.

Planned Action: none

(2.24) **Review: Why do authors not qualify the bias and variance of the estimator?**

Reply: Because we have organized the paper such that first the methods are presented, and in Section 6 error estimation is discussed. This seems justified to us, because a lot of the error propagation stuff can be treated in parallel for all the estimators, and touching this issue here would lead to redundancies which would make the paper even longer. Both bias and variance of the estimates depend on many more choices than the estimator alone.

Planned Action: none

(2.25) **Review: Why the reader has to wait until section 6, just to see the variance alone of the estimator.**

Reply: An estimator does not have a variance, only the estimate has one. There are a lot more sources which contribute to the variance of the estimate than measurement noise. Making an exception for this particular source of variance does not seem adequate to us.

Planned Action: none

(2.26) **Review: Page 7, Eq. (4). This is the worst part of the paper. Equation (4) is the OE estimator where S_a^{-1} has been substituted with R. In force of this unjustified and adhoc substitution, authors claim that the estimator (4) becomes more flexible and powerful than the OE shown in Eq. (6).**

Reply: We do not make such a statement.

Planned Action: none

(2.27) **Review: Also, in this case the variance of the estimator has been presented to the reader in instalments; first Eq. (7) and then an incredible jump to go to Eq. (18).**

Reply: We find it quite natural to first present the methods and then discuss the error sources. This seems particularly adequate to us since Eq. 18 represents only one component (often not even the leading one!) of the random error.

Planned Action: none

(2.28) **Review: In addition, (a) The bias of the estimator is not qualified/assessed/quantified in any part of the document**

Reply: The bias caused by the regularization is only one component of the total bias. We do not see any good reason to give it an extra treatment by discussing it in Section 4 while all other bias-generating errors are discussed in Section 6. This would disrupt the logical structure of the paper and may even lead the readers astray because they may think that the bias caused by the regularization term is always the most important one.

Planned Action: We will rewrite Section 6.4.5 and will discuss the bias-generating properties of the retrieval approaches there.

(2.29) **Review: b. What is the reason to change S_a^{-1} with R? What are the expected improvements?**

Reply: We do not claim in this paper that there are improvements. We simply want to systematize existing retrieval methods by presenting them in a common framework and notation.

Planned Action: none

(2.30) **Review: c. Why has the TikhonovTwomey regularization γ-parameter disappeared? That is why not γR?**

Reply: Thanks for spotting!

Planned Action: The equation will be corrected. The text above the equation will be changed to: '[... first order differences matrix,] and γ a scaling parameter to control the strength of the regularization".

(2.31) **Review: d. What's the role of x_a, and why not x_0 as in Eq. (3)?**

Reply: It makes a difference with respect to what the solution is smoothed. The solution of Eq (3) (in the linear case or after iteration in a well-behaved case) does not depend on x_0. Thus x_0 can be freely chosen. The solution of Eq (4) does depend on x_a, because the smoothing operator will not smooth the profile but the difference between the profile and the a priori. Thus, x_a cannot be freely chosen.

Planned Action: We will point this issue out in the text

(2.32) **Review: e. With R set to any of the suggested matrices, 012 order difference matrices, Eq. (4) is dimensionally inconsistent. The authors seek a protocolindependent of constraints and other assumptions, but they propose to use an estimator, which is dimensionally inconsistent and depending on the units used for the state vector. In which way do they achieve dimensional consistency between the two terms in the squared brackets?**

Reply: First of all, we do not propose anything, but we describe methods which are actually in use. And back to the question: By an appropriate definition of γ (which has admittedly been missing in the discussion paper) dimensions can easily be included.

Planned Action: We will define γ in the text and mention its units.

(2.33) **Review: It would be much fairer to say "Equation (4), as well as Eq. (3) (e.g. global fit), has been normally in use for the retrievals from satelliteborne limb sounding and occultation observations. It is here considered because still now many satellite processors rely on it. Or something similar. The description of the various estimators, LS, TT, OE should be as much as neutral and respond to the need to just explain their error characteristics.**

Reply: Eq (4) is the algebraic generalization. Both Tikhonov smoothing and optimal estimation are particular instanciations of it. We think that this is a fairly neutral way to present these methods. It shows how themethods are related.

Planned Action: none

(2.34) **Review: Page 7, line 30. What do you exactly mean with smoothed? What is a smoothed profile? How smoothing is quantified, and why this is a good property.**

Reply: A smoothed profile is a profile where the altitude-to-altitude differences of the profile values are reduced. The question why it is a good property is answered in the second part of the criticized sentence: "thus avoiding unphysical oscillations..."

Planned Action: We will modify the sentence as follows: "[...smoothed] in the sense of reduced altitude-to-altitude differences".

(2.35) **Review: In comparison to estimator (3), estimate (4) is biased and the bias structure is determined by R, which is grid dependent. So, how the estimated errors can be propagated according to CoA2? What is the solution proposed by authors: just forget about bias?**

Reply: If the retrieval is conceived as an estimate of the smoothed truth as discussed in Section 6.4.2 and if the measurement response as discussed in Section 6.4.5 is unity (as it typically is with first order differences Tikhonov regularization) then estimator (4) is bias-free. If $\mathbf{S}_\mathrm{a}$ does not equal the (typically unknown) $\langle \vec{x}_\mathrm{true} \rangle$, optimal estimation will have a bias. Thus, things are not as simple as they seem to be. We thus think that the bias discussion should not be touched upon passing in Section 3 but should be deferred to Section 6.4.5, where we have the content of Section 6.4.2 available, and where we can discuss the bias issue at more depth.

Planned Action: Section 6.4.5 will be rewritten to include the bias issue.

(2.36) **Review: Page 8. Eq (6). Now that the authors have invented R, they can say our estimator retrieves the OE estimate if we put $R = S_a^{-1}$, unbelievable!**

Reply: We find it quite natural that, when we generalize over formalisms and then specify again, we get the original specification back. We do not see what is wrong about this.

Planned Action: none

(2.37) **Review: By the way, to me, to $R = S_a^{-1}$, is the only possible choice, if we want to reach dimensional consistency.**

Reply: We disagree. With the correct units (which can be imported via γ), any $\mathbf{R}$ will be dimensionally consistent, regardless if it has a probabilistic interpretation or not.

Planned Action: none

(2.38) **Review: Page 8, paragraph beginning at line 8. This comment seems to stay here just to add some references.**

Reply: The fact that in the case of logarithmic retrievals the data characterization also refers to the logarithm of the state value is often overlooked and has already caused some confusion among data users. Thus, we find it appropriate to mention this issue.

Planned Action: none

(2.39) **Review: By the way, it is not appropriate for Eq. (6). This is a comment to be added soon after Eq. (5). It does not apply to Eq. (6), in fact, OE elegantly solve the problem of high dynamic range of the state vector, because S_a has the right dimension to properly scale the state vector. As shown in many papers, OE can be solved for the scaled variable $\tilde{x} = S_a^{-1/2}x$, which is equalized to a standardized variate, at each layer.**

Reply: We disagree. The caveat regarding the Gaussian probability density function is relevant only if the estimate is given a probabilistic interpretation, i.e., in the context of Eq. (6). And the suggested method using $\tilde{x}$ as a retrieval variable does not solve the problem that, for a variable which mostly has small values but a large natural variability (i.e. large x_a), the wings of a Gaussian penetrate wide into the negative. That is to say, optimal estimation assigns positive probability densities to negative temperatures or mixing ratios.

Planned Action: We will better highlight the problem of positive probability densities to negative temperatures or mixing ratios.

(2.40) **Review: Page 8, Eq. (7) and discussion after. Here it seems that an essential role in error estimation is played by the variance of the estimator alone, and the bias? Once again, how the bias of estimator (4) is qualified/assessed/quantified?**

Reply: Here we neither discuss the variance nor the bias. Both variance and bias include more than only noise and regularization, respectively. Thus, the discussion of both is deferred to Section 6.

Planned Action: The bias will be discussed in Section 6.4.5.

(2.41) **Review: Section 5. All is said in this section is today overcome**

by Simultaneous Retrieval. Section 5 is outofdate and should be totally removed.

Reply: A scientist trying to figure out what the total error budget of HALOE or SOFIE data is, is not much helped by this statement. And for, e.g., infrared spectroscopic instruments with 30-40 data products, represented at tens of altitudes each, and – depending on the instrument type – more than 1000 profiles per day with overlapping lines of sight, simultaneous retrieval of everything is still beyond reach. And if, e.g., spectroscopic data of one species are inconsistent in different parts of the spectrum, simultaneous retrieval can even be worse than a sequential approach.

Planned Action: none

(2.42) **Review: Section 5.4.5 Still Onion Peeling?**

Reply: As said above: the users of, say, HALOE or SOFIE data are not helped very much by saying "the data providers should have used another retrieval method."

Planned Action: none

(2.43) **Review: Section 5.4.6 See point above. I recommend a CoA0: Please forget about adhoc and nonoptimal methods!**

Reply:
1. It is the purpose of this paper to get error estimation for existing data sets under control. We are not proposing a data analysis scheme for a future instrument. The reviewer seems to have misunderstood the conditions of adequacy. They are not about retrieval schemes, but for error propagation schemes for given (not necessarily favoured or endorsed) retrieval schemes.
2. Optimal methods are optimal only if a real x_a and a real S_a are available. These are often not available, and many "optimal estimation" retrievals are non-optimal retrievals in disguise. Some of the instruments covered by our study have made measurements of some species for the first time. Where to get the prior and its statistics from in this case? And finally: Who says that the prior which was valid until yesterday is still valid today? Remember the turkey that came to the gate of the enclosure everyday at 9:00 expecting to be fed. This went well until Thanksgiving. But according to the rules of inductive inference, on which optimal estimation is based, the turkey behaved fully rational!
3. Forgetting methods not favoured by the reviewer clashes with comment 2.33, where we are requested to be neutral? (cf 2.33)
4.We understand that science is the generation and aggregation of knowledge. Based on this assumption it is unclear to us how forgetting anything should advance science.

Planned Action: none

(2.44) **Review: Sections 6.1 to 6.3 can be summarized under a very short section entitled "Instrument Noise and Forward Model bias"**

Reply: First, we have organized Section 6 by causes of the errors and not by random versus systematic errors. This is because the same source of error can show up as the one or the other, according to the retrieval scheme. And second, we do not see how this reorganization should make the section shorter.

Planned Action: none

(2.45) **Review: Section 6.4. Authors here simply miss the important point that the Averaging Kernels matrix, A, qualifies and serves to assess the bias error, at least the part coming from the background constraint. In fact, if we take expectations on both side of Eq. (25) all random components associated to the instruments are averaged to zero, and we remain with the expectation value, $E(\hat{x})$. Systematic component, originating from the forward model, can be dealt with appropriate transforms of the radiance vector, e.g., random projections.**

Reply: The bias caused by the regularization is dealt with in Section 6.4.5.

Planned Action: Section 6.4.5 will be expanded and restructured.

(2.46) **Review: Section 6.4.1. All the verbose premise of the paper points straight to this criticism of the smoothing error. However, the only thing which is fairly criticisable here is the word smoothing. In fact, smooth, smoother and similar terms should be banned from the context of error assessment and analysis. If Rodgers had said the retrieval can be regarded as a biased estimate of the true state, then everything would have gone to the right place. In effect, the smoothing error is the missing bias term to be added to the variance in order to have an estimate of the root mean square error, $E[(\hat{x}-x)^2]$. In principle, there is no need to interpolate/extrapolate to different grids a given state vector for the purpose of comparison. For visual inspection, one can just plot the given estimators and confidence intervals on the same plot, using the proper pressurealtitude grid. Why the quest of plotting differences?**

Reply: We do not criticize the smoothing error as such but we criticise that it may be included in the total error and will thus be inappropriately propagated for resampled profiles. We find the claim that interpolation is unnecessary somewhat odd. Science does not only consist of plotting data. Some more quantitative approaches are required. Time series at one altitude, when the original data have a varying altitude grid, quantitative profile comparison as suggested by Rodgers and Connor (J. Geophys. Res. 108(D3):4116, doi10.1029/2002JD002299, 2003) and many more scientific applications need interpolation of the data to a common grid.

Planned Action: none

(2.47) **Review: Pag. 27 and 28. Eq. s (28) and (29) can be left to more elaborated comparisons. There is no need to cover this aspect in the present paper.**

Reply: Here we agree with reviewer #1 who finds this section paticularly important. Furthermore, these sections are important to understand when regularization can cause a bias and when not. Equations 28 and 29 are essential for these sections.

Planned Action: none

(2.48) **Review: Pag. 28. Eq. (30). What do you mean "better resolved"? Please, quantify. The paper is aiming at providing recommendation, this cannot be given in terms of ambiguous qualitative terms.**

Reply: This statement refers to situations only where the contrast in the resolution is large. Thus, this statement does not depend on the particular definition of vertical resolution, any of the resolution concepts introduced in Section 6.4.3 will do.

Planned Action: We will move Section 6.4.3 before Section 6.4.1. Then we will have the definitions of altitude resolution available and will make reference to these definitions.

(2.49) **Review: Pag. 6.4.3 From section 6.4.3 on, until section 7, the paper appears to be unnecessary long.**

Reply: At many instances the reviewer criticizes that the various consequences of regularization are not sufficiently discussed, and here, where these issues are dealt with, the paper is criticized to be too long. We are confused.

Planned Action: none

(2.50) **Review: Section 7. As said at the beginning 18 recommendations are too many to be useful.**

Reply: As said before, we would have preferred less recommendations but condensing them makes them less specific, and finally we would end up with some vague truisms which would not be helpful at all.

Planned Action: none

(2.51) **Review: Table 1 and Table 2. I do not understand the scope of these two tables. If authors want to provide a list of official L2 data providers, the list is too long since it should show only Agencies. If the authors want to provide a list of the many scientists dealing with Satellite Data Processors, it is too short**

Reply: "official L2 providers" and "agencies" are no terms of scientific relevance. And including a "list of the many scientists dealing with Satellite Data Processors" would not be useful either. Our criterion is: we included data processors of which the data are distributed to the scientific community. We think that this is a sensible criterion.

Planned Action: none unless we are made aware of further data products that deserve to be included according to the criterion mentioned above.