Anonymous Referee #3

We want to thank referee 3 for commenting on our manuscript. His/her comments helped us to further improve our paper and triggered an interesting scientific discussion. For clarification, the comments 2a, 2b, 7a, 7b, 7d, 14 (see below) of this referee address one issue, which can be solved easily by specifying an incomplete statement in the abstract and in the introduction of the manuscript (as stated by referee 1 under minor comments). With this minor correction, we are convinced that all our statements are mathematically correct, that our paper is clearly structured. and that our approach to calculate total column averaging kernels has scientific relevance. The original comments of the referee are numbered, printed in blue and set in quotations marks.

I. Scientific significance:

1)

"The method claims to combine the numerical simplicity of a profile scaling retrieval with the numerical robustness of more sophisticated schemes. The Jacobian used in a profile scaling retrieval is only a vector, containing the sensitivities of the measured values with respect to the scaling factor. The method of the paper under review however, assumes availability of the full Jacobian matrix (i.e. altitude-resolved partial derivatives). It remains unclear, why, with the full Jacobian available, it should be disadvantageous to calculate the averaging kernel in the traditional way, e.g. either for a strongly regularized retrieval which leaves only one degree of freedom, or for an altitude-resolved profile retrieval from which finally the column is calculated. The "simplicity of a profile scaling retrieval" involves that this Jacobian is NOT available. To make a strong case that this method is actually useful and advantageous is crucial for this paper. Frankly speaking, I do not quite see what the problem is with the conventional approaches."

The referee overlooks the difficulty of calculating a Jacobian with respect to a profile scaling factor without referring to the availability of the full Jacobian. Even more, it is totally open in his/her comment how such a vector should be obtained. To clarify this, only in the case of a very simple radiative transfer model (e.g. transmission model for a pressure and temperature independent absorber), the derivative with respect to the scaling factor can be directly calculated. One may also calculate the scaling factor Jacobian with a numerical perturbation scheme using a standard radiative transfer model. But in this case, one suffers from finding an appropriate perturbation strength suited for all geometries and atmospheric situations. In principle, such an approach cannot be favorable for a robust retrieval approach. This is also the reason why the altitude resolved Jacobian is generally also available in 'simple' profile scaling algorithms. So, the statement of the review that "simplicity of a profile scaling retrieval involves that this Jacobian is NOT available" is not true. The simplicity of the profile scaling approach lies in the inversion procedure (i.e. a simple least squares approach can be used).

The gain in the numerical performance of our proposed method for profile scaling retrievals with respect to the traditional way of an altitude resolved profile retrieval is approximately equal to the number of vertical model layers n (we derive this estimation in our answer to comment 2 of referee 1). This is a significant improvement over the "traditional way" mentioned by the referee. For example, we showed in our practical investigation of the required vertical grid, which was rated by referee 1 as interesting and relevant contribution, that 20-40 vertical model layers are required to correctly calculate total column averaging kernels. Additionally, a sufficient number of vertical layers is also required to realistically represent atmospheric profiles during the retrieval.

Our paper is the first to come up with an approach to calculate the averaging kernel analytically for profile scaling algorithms without the need for simulation via a full profile retrieval or a perturbation method. We are convinced this is important for all scientists working with profile scaling algorithms, as now they are able to compute accurate averaging kernels in an efficient manner. For example, our approach could be useful for the Total Carbon Column Observing Network (TCCON) which at the moment supplies total column averaging kernels only for representative scenarios at two stations and than compiles those in a look up table for global comparison (please read comment 7 of referee 1). TCCON is an unique source for the validation of global models and satellite measurements and can be even more worthwhile when averaging kernels are available for each station and individual measurements. This is in particular important when a high accuracy is required (for example for XCO2 and XCH4), but it is also advantageous for the comparison with upcoming satellite missions like TROPOMI which will deliver total column averaging kernel for individual measurements.

II. Presentation and technical/scientific errors:

2 a)

"At many instances the language is sloppy, misleading, or even erroneous (semantically, not syntactically). Further, the paper is organized in a way that it is virtually ununderstandable until one arrives at Eq. (30). After having seen the concept introduced by Eq. (30), I had to start again to read the paper from the beginning of Section 2."

We understand the referee's point concerning the information of Eq. 30. It refers to an incomplete statement in the abstract and introduction. This can easily be improved by the following textual changes (as it was also suggested by referee 1 under minor comments (3)):

Abstract, L6-8:
from
"Formally, the proposed method is equivalent to Tikhonov regularization of the first kind with an infinite regularization strength**.**"
to
"Formally, the proposed method is equivalent to Tikhonov regularization of the first kind with an infinite regularization strength **and a vertical profile that is expressed relative to a reference profile**"

Introduction, p5003, L9-12:
from
"We will show that the proposed approach is equivalent to a profile retrieval using Tikhonov regularization of first order with an infinite regularization strength**.** However the approach preserves all advantages of a robust numerical implementation of the least-squares scaling approach."
to
"We will show that the proposed approach is equivalent to a profile retrieval using Tikhonov regularization of first order with an infinite regularization strength **and a vertical profile that is expressed relative to a reference profile.** However the approach preserves all advantages of a robust numerical implementation of the least-squares scaling approach."

Furthermore, we disagree on the referee's comment about the organization of the paper. In general, the paper provides first the theoretical framework (Sec. 2.1), which is elaborated in more detail in Sec 2.2. Here, the latter is needed to prove in Sec. 2.3 the statement that the column averaging kernel can be derived directly from the least squares scaling approach. Finally, we discuss two examples for its application in Sec. 3. In the sense, we approach the problem from the general mathematical concept to the specific application. So, it is obvious that the specific definition of the state vector as a relative profile is given in Eq. 30 and not earlier because the discussion before Eq. (30) provides a general framework of the paper. From Eq. 30 onward, the specific case of profile scaling is discussed. To avoid confusions, we specify the definition of state vector **x** after Eq (1) on p5004. We will change the statement from

'Here, the n-dimensional state vector **x** represents the vertical distribution of the trace gas.'
to
'The n-dimensional state vector **x** represents the vertical distribution of the trace gas. Here, a further specification of **x** or **y** is not required.'

2 b)

"Statements made earlier in the paper are simply incorrect without the specification given by Eq. (30) and mislead the reader."

We disagree.. The discussion until Eq. (30) does not require a specific definition of the state vector and is correct. It provides the general context.

3)

"Abstract and throughout the paper: The paper tries to maintain some kind of general applicability, but without specifying e.g. the measurement geometries under consideration, it is hard to judge if the statements are true or not. More specific and thus verifiable statements are needed instead of overgeneralized commonplaces for which almost always a counterexample can be found."

We agree that we should better specify the remote sensing problems for which our approach is applicable (please also see our answer to comment 7a of referee 1). We suggest the following changes to the introduction of the revised manuscript
e.g. at p5003 starting at L6:

from
"In this study, we present a concept for the retrieval of vertically-integrated column densities of atmospheric trace gases from remote sensing measurements which is based on the least-squares scaling of an reference profile but provides, in addition, an analytical expression for the column averaging kernel."
to
"In this study, we present a concept for the retrieval of vertically-integrated column densities of atmospheric trace gases from remote sensing measurements with typically one piece of information on the trace gas abundance that can be inferred from the measurement. The approach relies on fitting a least-squares scaling of a reference profile and provides, in addition, an analytical expression for the column averaging kernel. So, the retrieval allows one to determine the total column of a trace gas but not its vertical resolved distribution."

4)

"General: The term "column averaging kernel" is something different than the averaging kernel in the Rodgers book. The term is often used in the paper before it is defined on page 5007."

The referee overlooked the definition of the column averaging kernel at p5001, L28 – p5002 which defines the total column averaging kernel. To make this statement more clear, we propose the following changes:

from
"So, an analytical expression is given for the column averaging kernel which describes the sensitivity of the retrieved column with respect to changes of the true vertical trace gas distribution as function of altitude."
to
"So, an analytical expression is given for the column averaging kernel which describes the sensitivity of the retrieved column with respect to changes of the true vertical trace gas distribution as function of altitude and is thus defined by the corresponding derivative of the retrieved column"

5)

"p5001 l.22-24: This is grossly oversimplifying: The truth of this statement depends on the spectral resolution of the instrument, the measurement geometry, and the vertical grid chosen. This is a typical example where the desire to generalize leads to incorrect statements."

To clarify the statement we propose the following changes to the revised manuscript p5001,L22-24:

from
"A trace gas column retrieval represents a typical inversion problem of atmospheric remote sensing which is usually limited in its vertical sensitivity. This means an unregularized profile retrieval based on a standard least-squares fit yields a trace gas profile which is dominated by the contribution of measurement noise."
to
"A trace gas column retrieval represents a typical inversion problem of atmospheric remote sensing which is limited in its vertical sensitivity. This means an unregularized profile retrieval based on a standard least-squares fit applied to a remote sensing application with a sensitivity limited to the total column yields a trace gas profile which is dominated by the contribution of measurement noise."

6)

"p5002 l13: not quite clear what .normalized. means in this context."

We agree and propose the following change to the revised manuscript:
from
"More frequently used is a regularization approach which selects ad hoc the most representative vertical profile which is normalized to its vertical column amount."
to
"More frequently used is a regularization approach which selects ad hoc the most representative vertical profile which is divided by its vertical column amount. This normalization ensures that the retrieved scaling factor reflects the trace gas column amount."

7 a)

"p5003 9-12: 1st order Tikhonov is not equivalent with profile scaling. Scaling implies a multiplicative modification of the a priori profile while 1st order Tikhonov implies an additive modification (c.f. second term in Eq.(3)). For profiles with a large dynamic range, this difference can be dramatic. This difference and its implications are not at all mentioned in the paper until Eq. (30)."

We agree and propose to change the text as mentioned already in our reply to referee's comment 2b):

Introduction, p5003, L9-12:
from
" We will show that the proposed approach is equivalent to a profile retrieval using Tikhonov regularization of first order with an infinite regularization strength. However the approach preserves all advantages of a robust numerical implementation of the least-squares scaling approach."
to
" We will show that the proposed approach is equivalent to a profile retrieval using Tikhonov regularization of first order with an infinite regularization strength **and a vertical profile that is expressed relative to a reference profile.** However the approach preserves all advantages of a robust numerical implementation of the least-squares scaling approach."

7 b)

"On the contrary, the equivalence of profile scaling and Tikhonov 1st order with large lambda is explicitly stated on page 5010 l19/20. The only thing both these methods have in common is that there is only one degree of freedom left but this common feature means by no means that this is the same! This conflict is only remedied in Eq. (30) for a particular choice of representation of the state space"

We are convinced that the proposed changes in our reply to this referee's comment 2b) clarifies this point and that the statement in the introduction (p5003, L9-12) is now in agreement with the correct mathematical definition given on page 5010 L19-20.

7 c)

"but (a) here a kind of prior profile gotten rid of in Eq. (9) comes in through the back door again (via rho_ref),"

We disagree. **rho_ref** is the trace gas profile which is used for the scaling during the retrieval. This is very different than the prior profile $\mathbf{x_a}$ used in e.g. an Optimal Estimation approach (see the book of Rodgers, 2000). The prior profile in e.g. Optimal Estimation is used to fill up the null space of the matrix (**I-A**), where **I** is the identity matrix and **A** the averaging kernel (the definition is given in our equation (8)). Moreover, we showed at page 5015 L17-19 of the submitted manuscript that the a priori profile must be different to the reference profile, because by definition the reference profile does not have a null-space contribution in the case of a profile scaling retrieval. This is a result from equation (24) and is discussed accordingly in the manuscript.

7 d)

"(b) the whole discussion from p5004 to p5010 remains ununderstandable with respect to the equivalence of Tikhonov 1st order and scaling until Eq. (30) is presented. The reader is mislead over pages!"

This strong statement of the referee is wrong. Unfortunately, the referee assumes the state vector $\mathbf{x_a}$ to be identical to the profile **rho_ref** and thus the referee finds the discussion on p5004-5010 contradictory with respect to the rest of the paper. However, as discussed earlier, the paper does not make this assumption and so the referee's comment does not reflect the presented material. We think that the proposed reply to the referee's comment 2b clarifies this point sufficiently.

7 e)

"Also, l6/7 in the abstract are very misleading because the restriction is not mentioned."

This is corrected according to the comment 2b) of this referee.

8)

"p5004-5007: This is quite a broad description of the Tikhonov smoothing approach which does not add much to the existing (often not referenced) literature of the method and its application in remote sensing."

This statement is not correct. Section 2.1 on p5004-5007 gives the theoretical background of Tikhonov profile retrieval which is needed to understand the equivalence between the profile scaling approach and the Tikhonov regularization of first kind for infinite regularization strength. All relevant publications are cited. Citations regarding the application of this method in the field of atmospheric remote sensing are given in the introduction.

9)

"p5004 l5 and Eq (1): e_y are not .error bounds., but as used in Eq (1) these are the actual errors."

The referee's comment is correct. We propose to change the sentence from
"... forward model **F** which describes the measurement within the bounds of a spectral error $\mathbf{e}_y$, ..."
to
"... forward model **F** which describes the measurement within the spectral error $\mathbf{e}_y$, namely..."

10)

"p5004 l15: Here the use of .error bounds. is not consistent with Eq. (1)."

The referee's comment is correct. We propose to change the sentence:

from
"The retrieval of a trace gas abundance involves finding a state vector **x** which reproduces the measurement y within its error bounds via the forward model **F**"
to
"The retrieval of a trace gas abundance involves finding a state vector **x** which reproduces the measurement y within its error via the forward model **F**."

11)

"p5004 Eq 3: The Se-1 and the Ln-p matrices must not appear inside the norm signs."

We disagree. The measurement error covariance and the regularization matrix must be inside the norm. This is a common nomenclature for example used in the book of C. Hansen (1997) mentioned in the following answer to comment 12.

12)

"p5005 l3: I have not found the term "regularization matrix" in the Rodgers book."

The term "regularization matrix" is common and for example used in the well known book

"Hansen.C, "Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion, SIAM, 1997, page 73". We will change the citation accordingly.

13)

"p5006 l6: "contributions of the null space": I quote the Rodgers book, p 49 second paragraph: "This component of the error budget was described incorrectly as "null space error" in Rodgers (1990). The term null space error should properly be used to describe the contribution to the error budget from those components of the state which lie on the null space of K, and are consequently not seen by the retrieval...". Since it is, in the general case, not even clear that K has a null space, and since (I-A)x_a contributes depending on how large lambda is chosen, usually also to quantities not affected by the null space, I consider this terminology here and later in the manuscript) as inappropriate."

The referee is here pointing to the error term that Rodgers is denoting in his book as "smoothing error". We will rename the term "null space error" to "smoothing error" according to the nomenclature of the Rodgers book throughout the paper. Furthermore to prevent confusion, we change our definition of that term on p5006, L4-8 of the submitted manuscript:

from
"The contribution $\mathbf{A}\mathbf{x}_{true}$ is the part of the solution which can be determined from the measurement sensitivity, whereas $(\mathbf{I}-\mathbf{A})\mathbf{x}_a$ is the null space contribution of the solution using a priori knowledge about the atmospheric state. In other words, the measurement is effectively not sensitive to contributions of the null space."
to
"The contribution $\mathbf{A}\mathbf{x}_{true}$ is the part of the solution which is determined from the measurement sensitivity, whereas $(\mathbf{I}-\mathbf{A})\mathbf{x}_a$ is the part of the solution that comes from the a priori profile $\mathbf{x}_a$. In particular the null space of the matrix $(\mathbf{I}-\mathbf{A})$ describes all contributions of the state space were the measurement and by that the retrieval is effectively not sensitive to. Hence, this null space is known as the effective null space of the retrieval. The term $(\mathbf{I}-\mathbf{A})\mathbf{x}_a$ describes how the effective null space of the retrieval is filled up with the information taken from the a priori profile $\mathbf{x}_a$. It is the effective null space contribution of the a priori profile $\mathbf{x}_a$ and also known as smoothing error of the retrieval."

14)

"p5006 l9 and Eq. (10): When regularization is chosen so strong that the Tikhonov retrieval actually mimics a column retrieval (i.e. a retrieval with only one degree of freedom), then the result according to Eq. (10) will be an altitude-constant profile (because the regularization forces differences between x_n and x_n-1 towards zero. I found it bold to stretch linear theory so far that it is assumed that the Jacobians are the same for a straight line and a realistic profile, before I learned in Eq 30 how the x-vector is re-interpreted. In other words: Without having seen Eq. (30), it is not possible to see anything useful w.r.t. column retrievals in Eq. (9)."

We disagree with the referee. As already mentioned above, the discussion in Sec. 2,1 provides the overall theoretical concept of the paper and is essential for the rest of the manuscript. In this section, we do not use a specific representation for the state vector before equation (30) and in Sec. 2.1 all equations are correct and valid for all representations of the state vector.

15)

"p5009: Here the meaning of the different terms are discussed but it would be helpful to finish this section with explicitly writing out the transformed retrieval equation and the transformed equation used to calculate the averaging kernel. Otherwise the reader is left with backsubstitution of all these terms. Further, it is not quite clear what is gained in comparison to the traditional method (except better insight into the anatomy of the Tikhonov method)."

We do not agree. We intentionally decided to derive an analytical form of the gain matrix **G** which allows to calculate the averaging kernel via a simple matrix multiplication **A**=**GK**. This puts the focus on what we need in the further discussion. The back substitution mentioned by the referee is trivial and would only result in a huge expression that is superfluous. We do not understand the referee's point of view, who asked us to omit essential material as given in Sec. 2.1 and at the same time suggests to work out a straight forward back substitution here.

16)

"p5006 "integration": is the state space represented by continuous functions, or should it read "summation"?"

The operator C represents an altitude integration of the state **x**. In this context, the summation is a technical term and does not explain the definition of C in the text below Eq. (11).

17)

"p5009: The term .degrees of freedom for signal. is inferred on p29-31 of the Rodgers book explicitly for maximum a posteriori retrievals. Particularly, his Eqs 2.51-2.53 make use of the fact that x_a is actually the mean of the true state. For general regularized retrievals this is nottrue, so the term .degrees of freedom for signal. should not be used. The commonly used more generic term for the latter case is to my knowledge. "degrees of freedom of the retrieval" or "degrees of freedom of the retrieved profile"."

We agree and accordingly will correct this in the revised manuscript at p5009, l15:

from
"... known as the degrees of freedom of signal (dofs)."
to
"... known as the degrees of freedom of the retrieval (dofs)."

18)

"p5010 l1-10: This is neither new nor relevant to this paper; it is not used in the following."

We agree with the referee and accordingly will remove L1-10 of p5010 from the manuscript.

19)

"p5010 Eq. (27): This equation is involves an undefined sum because of non-matching matrix dimensions (the sum is defined only for the squares of these matrices)."

This equation will be removed from the new manuscript (please see comment 18).

20)

"p5012 l17: While the paper by von Clarmann and Grabowski might be relevant to the paper under review in a more general sense, because these authors use excessive (in their case: blockwise)

Tikhonov smoothing to emulate (in their case: partial) column retrievals, as in the paper under review, this paper is quoted somehow out of context: The Tikhonov retrievals in their paper are used in the first place to find out how a reasonable altitude grid for an unregularized retrieval might look like, and everything else is linear transformation. Availability of intra-layer averaging kernels is only a sideproduct of their approach."

We agree and accordingly will introduce an additional paragraph in the introduction of the revised manuscript to pose this citation right starting at page p5003 L5:

"Another method to calculate total column averaging kernels for the profile scaling approach is the simulation via a particular form of a Tikhonov profile retrievals as it is done for example by von Clarmann and Grabowski (2007) and Sussmann and Borsdorff (2007). The advantage of this approach is that the total column averaging kernel can be calculated analytically but it is only valid in theory since an infinite regularization strength must be chosen for it. Hence, for practical applications this can only be approximated and a careful tuning is necessary to avoid numerical instabilities. Furthermore this simulation requires a full profile retrieval with n layers even though only one scaling parameter is estimated. Therefore, an operational implementation of this approach would negate the computational advantage of the original profile scaling method."