Anonymous Referee #1

We thank referee 1 for commenting on our paper and by that helping to improve our manuscript. Our replies to his/her comments are given below. The original comments of the referee are numbered, printed in blue and set in quotations marks.

I. Major comments:

1 a)

"The authors present some considerations concerning scaling retrievals and the calculation of associated averaging kernels. The topic is carefully developed, but I have to admit that in my opinion the main conclusions can be derived in a much more straightforward manner without invoking the similarity between an L1 Tikhonov and the scaling approach."

We think that the chosen way to present the subject is indispensable for the paper. The referee suggests that the findings can be derived in a more straightforward manner, but leaves it entirely open what his/her suggestions are. This makes it not possible for us to react adequately to the reviews comment.
One of the main findings of the manuscript is the proposed approach for calculating total column averaging kernels for profile scaling retrieval algorithms. Even though column averaging kernels can be calculated in a straightforward manner using a regularized profile retrieval, a corresponding approach for a profile scaling algorithms is not obvious. Thus from this perspective, it is appealing to interpret the profile scaling approach first as a particular form of a regularized profile retrieval and than adopt from that expressions to calculate the column averaging kernel analytically. The mathematics of the manuscript is presented in a rigorous way and proofs this relation between the two inversion approaches.

1 b)

"I believe that the resulting relations are already in common use in codes (and associated tools for the characterisation of retrieval results), e.g. for the analysis of ground-based spectra. It is appropriate that the authors stress the importance of (column) averaging kernels for ensuring a proper use of the data, e.g. for a comparison with a model. The practical investigation of the required vertical grid width performed by the authors is an interesting and relevant contribution."

To our best knowledge, the results of our paper have not been published before. It describes a novel approach to calculate the averaging kernel analytically for profile scaling algorithms without the need for the simulation via a full profile retrieval or the application of a perturbation method. In our manuscript, we discussed these different approaches to calculate the column averaging kernels as they are presented in literature. In particular, the two methods, published so far, to calculate column averaging kernel for profile scaling retrievals are cited and discussed in our paper, (p5002, L28 - p5003, L5), (p5012 L13-19) which are:

1) Buchwitz et al. (2004) and Gloudemans et al. (2008) estimate the column averaging kernel by numerical perturbations of the retrieved column by means of changes in the true profile using simulated measurements. For the n vertical profile layers, the actual retrieval and also forward calculation must be repeated n times which is not efficient. Furthermore, the perturbation

strength must be chosen carefully to avoid numerical instabilities.

2) The method suggested by von Clarmann and Grabowski (2007) and Sussmann and Borsdorff (2007) is based on a profile retrieval approach represented on n vertical layers using Tikhonov regularization. This approach requires the numerical effort of a full profile retrieval estimating n parameters although the physical problem has, by definition, only one degree of freedom. In addition, a careful tuning of the regularization strength is needed which can lead to additional numerical complexity and may affect the robustness of the approach.

One aim of our paper is to introduce an alternative method, which allows to use the frequently employed profile scaling approach with minor modification to calculate the column averaging kernel analytically. (see the recipe on p5011, L4 – p5012, L7). To our knowledge, this approach is not published so far. In case referee 1 is aware of such a publication we would appreciate to learn more about that.

2)

"Because the authors claim that the algorithm is efficient and well-suited for operational analysis of satellite data, it would be desirable to include a comparison of the computational efforts required for (1) the simple scaling retrieval without generation of column averaging kernels (2) the proposed processing scheme: scaling retrieval including calculation of column averaging kernels (3) the constrained profile retrieval + generation of required diagnostics (averaging kernels). If the major computational effort is spent for the forward calculation and the generation of the required derivatives, I would not expect that the most flexible retrieval setup (3) would significantly increase the computational cost wrt approach (2). If the generation of altitude-resolved Jacobeans is considerable, an optimized approach (1) would be significantly more effective than (2) and (3), but would require additional consideration of providing a lookup table for the column averaging kernels as function of relevant parameters."

We agree with the referee that an estimate of the computational cost will strengthen the paper and its findings. However, to include the forward model in this comparison precludes an objective statement (please see also our answer to comment 1 of referee 3). A numerical expensive forward model will always outbalance the trade-off in the numerical effort of different inversion approaches, whereas for highly efficient forward models the numerical performance of the inversion becomes important, particular in the context of an operational data processing. We propose to change the manuscript accordingly to the discussion below.

For this purpose, we are convinced that case (1) mentioned by the referee does not provide further insight in the numerical performance. Commonly, the calculation of the total column averaging kernel relies on a vertically resolved Jacobian. An exception may be the numerical perturbation approach, which is not desirable because of the arguments given above, and may only work in case of very simple forward models which depend directly on the total column of a trace gas.
Due to that, we focus the discussion on the numerical cost of the different inversion approaches based on the number of the most costly numerical operations. To compare case (2) and (3), we consider the number of multiplications which have to be performed for the different inversions in one iteration step. Faster operations like summations are ignored. We assume dim($\mathbf{y}$)=m and dim($\mathbf{x}$) = n. For simplicity, we assume that a matrix multiplication like $\mathbf{K}^T\mathbf{S}_e^{-1}$ requires nm^2 operations and the inversion of a nxn matrix n^3. We ignore the inversion of the matrix $\mathbf{S}_e$ because for uncorrelated measurements noise this is trivial.

(case 3: constrained profile retrieval + calculation of the averaging kernel):

a) calculating the gain matrix **G** needs ($\mathbf{nm^2 + 2n^2m + n^3}$) operations.
b) The retrieval result **x=Gy** needs (**mn**) operations
c) The averaging kernel **A=GK** needs ($\mathbf{n^2m}$) operations

The total number of operations is ($\mathbf{mn + nm^2 + 3n^2m + n^3}$).
Assuming m>>n we approximate this number to be ($\mathbf{nm^2}$)
and in the case of m==n to (**5n3**)

(case 2: profile scaling + our method to calculate the total column averaging kernel)
a) calculating the gain vector $\mathbf{g_{lsq}}$ needs only ($\mathbf{m^2 + 2m + 1}$) operations
(because the vector $\mathbf{K_{lsq}}$ has the dimension m.)
b) The retrieval result $\mathbf{x_{lsq}=g_{lsq}*y}$ needs (**m**) operations
c) The total column averaging kernel $\mathbf{a=g_{lsq}*K}$ needs (**nm**) operations

The total number of operations is ($\mathbf{m^2 + 3m + 1 + nm}$).
For m>>n, this can be approximated to be ($\mathbf{m^2}$)
and in the case of m==n to ($\mathbf{2n^2}$)

Hence, this rough estimate shows that our proposed inversion approach is a factor of n faster than the established approach. In particular, for the applications presented in section 3 this means a speed up of about 20-40 for the inversion. But for smaller retrieval windows with a number of spectral points comparable to the number of profile layers this speed up would even further increase to 2.5n.

II. Minor comments / typos:

3)

"Abstract: "The proposed method is equivalent to Tikhonov reg of the first kind ...". This is only valid if the state vector is used as given in Eq. 30. If the state vector contains e.g. concentrations or mixing ratios the equivalence does not hold. Please state more precisely."

We agree with the referee. For example, a precise statement is given in the summary of the submitted manuscript p5017, Ll12-14. Accordingly we suggest the following changes:

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."

4)

"Page 5001: "this measurements" -> these measurements"

We agree and propose to correct this in the revised manuscript.

5)

"Page 5002: "In an ideal case, the column averaging kernel is constant ..." Ideally, the column sensitivity should not only be constant, but unity at all altitudes. "

We agree, but also see that this statement depends on the particular representation of the profile (e.g. a representation in partial columns or column averaged density). Thus, we propose the following changes at p5002, L3-5:
from
"In an ideal case, the column averaging kernel is constant with altitude, representing a retrieval product which corresponds to the geometrical integration of the true trace gas profile."
to
"In an ideal case, the column averaging kernel represents a geometrical integration of the true trace gas profile."

6)

"Page 5002: .Despite its theoretical advantages, only a few retrieval algorithms use this approach ..." - Not clear to me which aspects of the retrieval the authors refer to."

We agree and propose to change the sentence (p5002,L8-11):
from
"Despite its theoretical advantages, only a few retrieval algorithms use this approach, notably RemoteC (Butz et al., 2010) and ACOS (O'Dell et al., 2012). The reason for this could be the numerical complexity of the approach and its dependence on the particular regularization of the cost function."
to
"Despite its theoretical advantages, only a few retrieval algorithms derive trace gas columns and the according total column averaging kernels via an profile retrieval, notably RemoteC (Butz et al., 2010) and ACOS (O'Dell et al., 2012). The reason for this could be the numerical complexity of the approach and its dependence on the particular regularization of the cost function."

7 a)

"Page 5002: "[use of scaling retrieval within TCCON]... . But its main drawback is the lack of the corresponding column averaging kernel." - The main drawback of a scaling retrieval lies in the fact that the total column sensitivity shows larger deviation from ideal behavior than a profile retrieval. In general, a scaling retrieval provides a poorer reconstruction of the actual column, but it is numerically more efficient than a profile retrieval, especially if the column averaging kernel is not calculated for each individual retrieval."

We do not agree with the referee. His/her conclusion depends critically on the sensitivity of the measurement. To clarify our intention, we propose a statement in the introduction of the revised manuscript that we consider measurements with a low information content (dofs close to 1) which have a sensitivity mainly limited to the total column of a trace gas e.g. at p5003 starting at L6:

The statement is changed 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".

For these particular cases, we think that the a full profile retrieval has no additive value with respect to a profile scaling approach, provided that the column averaging kernel is calculated in our proposed manner.

7 b)

"The calculation of the column averaging kernel requires calculation of altitude-resolved Jacobeans, which is expensive. For this reason, TCCON uses a set of reference kernels (as function of solar elevation). A similar approach might be feasible for operational analysis of satellite data (lookup tables containing column averaging kernels as function of relevant parameters, e.g. ground albedo, cloud top height and cloud fraction)."

The column averaging kernel depends on many atmospheric parameters and solar and viewing geometry. This is why it is important, in particular for future satellite missions, to supply total column averaging kernel for each individual remote sensing measurement. For example, currently TCCON supplies total column averaging kernels calculated for only two stations (Lamont, Oklahoma and Park Falls, Wisconsin) discretized for 13 different solar zenith angles. For the remaining 13 global stations TCCON proposes to use the Lamont averaging kernels when comparing against this data set. This Information can be found on the TCCON Wiki under "Auxiliary Data, A Priori Profiles and Column Averaging Kernels for the (obsolete) GGG2009 Data". The TCCON dataset is a valuable source for the validation of global model calculations and satellite data and would be even more worthwhile when total column averaging kernels are supplied for each TCCON station and for each individual measurement. This is in particular important when

a high accuracy is required (for example for XCO2 and XCH4). Enlarging the lookup table may improve the accuracy but online calculation is favorable, to our opinion. In that sense, a lookup table approach seams not appealing to us.

8)

"Page 5003: "insides" -> insights"

Will be corrected in the revised manuscript.

9)

"Page 5004, Eq. 3: The first term in the cost function is incorrect (and would not be unitfree for a spectrum in radiance units). Please note: If you allow for a state vector which bears units (.the particular form of fk depends on the units of the state vector. . .., page 5007), then the entries of the L matrix would bear units, too (whereas Eq. 4 suggests that the L-matrix is unit-free)."

Will be corrected in the revised manuscript. In particular, the symbol **Se-1** is replaced by Se-1/2 to correct the first term and to make it unit less. Additionally we added the missing indicators for the L2 norm (a subscript 2) in this equation and at p5004, L22. Indeed, The **L** matrix can be kept unit less by assuming that the regularization parameter lambda is not unit less.

10)

"Page 5007, before Eq. 14: wrong symbol in the second equation"

Will be corrected in the revised manuscript. The brackets around **x** will be removed.

11)

"Page 5014: "faction" -> fraction"

Will be corrected two times in the revised manuscript (p5013 and p5014)