Comments on the revised paper "A practical method to remove a priori information from lidar optimal estimation method retrievals" by Ali Jalali et al

I am writing in response to the author's last updated version of their paper and their response to the raised comments. Although having ample time the authors have failed to respond to some of the serious concerns raised. Unfortunately, they have ignored some of the raised questions and their response to some of the other comments only raise more questions on the validity of the method presented in the paper. Although the equations provided in section 3.2 are correct they are irrelevant to this paper and cannot be used for the proposed methodology. This has been discussed in detail in the next paragraphs. After reading their responses and the revised paper, I am not convinced that the proposed method by the authors is scientifically correct.

Comments on section 3.2 of the revised paper:

I fail to see how section 3.2 of the present paper is relevant to the methodology of this paper. Surely, in section 10.3 of Rogers, the maximum likelihood retrieval is proposed (based on transformation to a coarser grid). But, unlike the proposed likelihood solution in the present paper, the maximum likelihood solution is evaluated as:

$$\hat{\boldsymbol{z}} = [(\boldsymbol{KW})^{\boldsymbol{T}}\boldsymbol{S}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}}\,\boldsymbol{KW}]^{-\boldsymbol{1}}(\boldsymbol{KW})^{\boldsymbol{T}}\boldsymbol{S}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}}\boldsymbol{y} = \boldsymbol{G}_{\boldsymbol{ML}}\boldsymbol{y}, \tag{1}$$

$$\hat{\mathbf{x}} - \mathbf{x} = (\mathbf{W}\mathbf{G}_{\mathbf{ML}}\mathbf{K} - \mathbf{I})\mathbf{x} + \mathbf{W}\mathbf{G}_{\mathbf{ML}}\boldsymbol{\varepsilon}. \tag{2}$$

Where $\hat{\boldsymbol{z}}$, is the reduced presentation of the retrieval, $\mathbf{W}\mathbf{G}_{\mathbf{ML}}\mathbf{K} - \mathbf{I}$ is the smoothing error, and $\boldsymbol{A} = \mathbf{W}\mathbf{G}_{\mathbf{ML}}\mathbf{K}$ is the averaging kernel which cannot be the unit matrix (unlike what is presented in the paper). More details are provided in Rogers section 10.2 and 10.3. Therefore, unlike what is proposed in the revised paper one cannot just change the retrieval grid to the coarser grid and make the a priori covariance equal to zero and rerun the OEM to get the maximum likelihood solution. The reason is that, first: the current forward model is not linear, and a non-linear approach is needed to calculate the retrieval (either Newtonian or Levenberg-Marquardt approach). Later when using the nonlinear method, the Jacobian matrix should be transformed as: $\boldsymbol{K}' = \boldsymbol{KW}$ for all the iterations. Such a transformation matrix has neither been introduced nor discussed in the paper and the Jacobian as is shown in the paper is evaluated in the fine grid. Secondly, an important step after calculating the retrieval is missing. The retrieved quantity needs to transform back to the fine grid as shown in Eq. 2, which in this case as mentioned before will result in a non-unitary averaging kernel. Finally, the representation error $\boldsymbol{\varepsilon}_{\boldsymbol{r}} = \boldsymbol{K}(\boldsymbol{I} - \boldsymbol{WW}*)\boldsymbol{x}$ should be calculated as an additional error. To make this point clear, as the OEM fine grid retrieval and the proposed reduced presentation of the retrieval are in two different grids, to compare these two results, the discussed transformation presented above is needed (For more details refer to Rodgers section 10.2 and 10.3).

Furthermore, the general form of the maximum likelihood which is presented in this section is an irrelevant topic to the paper. The general form for the cost function can be written as:

$$Cost = \left(\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})\right)^{\boldsymbol{T}} \boldsymbol{S}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}} \left(\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})\right) + \text{Regularization term} \tag{3}$$

The first term of the above equation surely is the maximum likelihood solution and the second term can be any regularization term (a priori constrain or Tikhonov regularization). Thus, it is trivial that omitting the second term of the equation will result in a maximum likelihood solution. However, in ill-posed problems it is necessary to use a regularization term. As mentioned in Rodgers (section 10.4.2) if a maximum a posteriori method is given (in which a priori is used), it is possible to remove the effect of the a priori without rerunning the retrieval to obtain a maximum likelihood solution of the form:

$$\hat{\mathbf{x}}_{\boldsymbol{Ml}} = \left(\widehat{\boldsymbol{S}}^{-\boldsymbol{1}} - \boldsymbol{S}_{\boldsymbol{a}}^{-\boldsymbol{1}}\right)^{-\boldsymbol{1}}\left[\widehat{\boldsymbol{S}}^{-\boldsymbol{1}}\hat{\mathbf{x}}_{\boldsymbol{MAP}} - \boldsymbol{S}_{\boldsymbol{a}}^{-\boldsymbol{1}}\boldsymbol{x}_{\boldsymbol{a}}\right]. \tag{4}$$

However, this solution is only valid if $\mathbf{K}_{\mathbf{l}}^{\boldsymbol{T}}\boldsymbol{S}_{\boldsymbol{\varepsilon}}^{-\boldsymbol{1}}\boldsymbol{K}_{\boldsymbol{l}} = \widehat{\boldsymbol{S}}^{-\boldsymbol{1}} - \boldsymbol{S}_{\boldsymbol{a}}^{-\boldsymbol{1}}$ isn't singular. Thus, as suggested in both Rogers and von Clarmann and Grabowski (2007) a hard constraint should be used (a coarse grid representation of the MAP retrieval is needed). As mentioned in Eq. 8 in Von Clarmann and Grabowski (2007), an interpolation matrix (**W**) can be used to transfer from the fine grid to the coarse grid (which provides a constraint), and the resulting averaging kernel is shown in Eq. 15 of the mentioned paper. An appropriate representation of the retrieval in the coarse grid will result in an averaging kernel equals to unity if non-trivial solution for the following equation can be found:

$$\boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{R}\boldsymbol{W} = 0 \tag{5}$$

Which is Eq. 19 of Von Clarmann paper.

In section 3.1 of Von Clarmann and Grabowski (2007), at first only the resampling constraint is used (using Eq. 21); however, it is stated that: "Resampling of the constrained oversampled fine-grid retrieval, however, degrades the profile and further reduces the information below dg_fretrieval. This follows from Eq. (15), because Eq. (19) is not usually satisfied for arbitrarily chosen W and R matrices. This leads to k×k averaging kernels unequal unity, which is equivalent to less than k degrees of freedom. To compensate for this additional loss of information, it seems necessary to first remove the a priori information in the fine-grid retrieval. This, however, is not possible in most cases, since the unconstrained n-grid solution suffers from ill-posedness. Instead, we search for a new constraint R_0, which in effect is equivalent to the resampling of the retrieval to the appropriate coarse grid, i.e. which fulfills Eq. (19)." Thus, they used Eq. 10.48 of Rodgers such that a pair of **R**' and **W**' was presented to reserve the information. The blocks of the **R**' matrix is based on Tikhonov regularization terms (equations 29 and 30). The triangular representation in section 3.2 has different approach for the blocks in **R**'; however, is still based on Tikhonov regularization (equations 37 and 38). Both triangular and staircase representations have the same approach for the resampling process and thus, I fail to understand why the authors of the present paper are insisting to claim that they are using the triangular representation when they only use the resampling part.

In summary, in Von Clarmann and Grabowski (2007) both resampling (via interpolation) and re-regularization (via an adoption of Tikhonov regularization) have been applied to retrieve the new profiles. Resampling without re-regularization will degrade the degree of freedom and re-running the OEM as maximum likelihood in a way suggested by the authors is not correct. Moreover, if only the re-sampling is used the maximum meaningful height decreases and even in the case of using both resampling and re-regularization the maximum meaningful height cannot exceed the OEM result. Even if the authors managed to resample and then use the correct form of the maximum likelihood (hard constraint), the maximum acceptable height cannot exceed the OEM (soft constraint) cut-off height. The proposed cut-off in the present paper based on 60% uncertainty of retrieval is rather a wrong claim. Instead, as also mentioned in other previous comments, the degree of freedom should be used to find the maximum acceptable height which certainly in the case of presented paper will be lower than the OEM acceptable height.

To sum-up, the use of this method to retrieve temperature and water vapour when it results in high uncertainty, low vertical resolution and certainly lower acceptable height is not useful (when plenty of reliable climatologies are available). I strongly believe that the proposed method with such high uncertainties is not suitable for trend analysis purposes.

The main concerns regarding this paper that the authors have failed to respond to, are:

According to the manuscript the output of OEM from the first iteration is used for the second iteration of the coarse grid retrieval. Now it is claimed that OEM is running completely in the fine grid which means that the optimal solution has already been obtained. Therefore, running the OEM again without using the a priori cannot minimize the effect of the a priori as claimed by the authors since it is fully implemented in the first run. Again, no mathematical proof is provided to show that how the effect of the a priori is minimized.

The authors have failed to respond to the comment of choosing a cut-off of 0.9 instead of 0.8 for the temperature profile. Also, the authors response for choosing a cut-off of 0.9 instead of 0.8 for water vapor mixing ratio retrievals is contradictory. The cut-off criteria suggested by the authors in their final response, is considering the SNR measurements which was not mentioned before. "We have found that a measurement response of 0.8 corresponds to a SNR of close to 1 or less than 1, which is above the cutoffs of what we typically use for the traditional water vapour method and where we would no longer consider the measurements meaningful." This response contradicts the manuscript of the initial OEM water vapor paper published by Sica and Haefele. Here, I am quoting the Sica and Hafele (2016) "One advantage of the OEM is the ability to determine the relative contribution of the a priori information relative to the contribution of the measurements. We used this information in the retrieval of Rayleigh-scatter lidar temperature to determine a height below which the retrieval was primarily due to the measurement and not the a priori. **We found that the well-known criteria that the sum of the averaging kernels (e.g., the measurement response) exceeding 0.8 was consistent with the maximum height found from the trace of the averaging kernel matrix, that is the degrees of freedom. Both criteria are used in this study. We found in clear conditions that the measurement response function is in excellent agreement with the signal-to-noise, SNR of the photocount measurements.** In cloudy conditions, particularly during daytime, the measurement response often overestimates the height where the measurements have reasonable signal-to-noise levels. For cases such as the cloudy daytime case shown below, the more conservative height based on the degrees of freedom is used to specify the height at which the retrieval is primarily due to the measurements and not the choice of a priori". Thus, according to this statement for clear conditions the degree of freedom is consistent with the cut-off of 0.8, and the response function is in excellent agreement with the SNR. This means that the SNR cannot be taken as close and lower than one. For the cloudy conditions the conservative height based in degree of freedom was chosen.

Another concern is with the authors claim of reaching higher altitudes in their proposed method. As is quoted in their response "We think there has been some confusion as to what we mean by increasing final retrieval altitude. It is correct that it is not possible to add information to the retrieval and therefore the entire retrieval cannot go beyond the last point on the fine grid retrieval which we see it does not in the coarse grid averaging kernels. In this case we mean, **that the coarse grid increases the altitude at which we consider the retrieval to be meaningful.** We have also clarified this in our response to Kgaran". This statement is incorrect since in OEM the degree of freedom is in direct correlation with the maximum acceptable height. In OEM the degree of freedom is used to determine the acceptable height of the retrieval which in many cases is consistent with the cut-off height of 0.8 as was mentioned by Sica and Haefele 2016. Regarding the degree of freedom, Clarmann and Grabowski (2007) have stated in their abstract that: "Since **regridding implies further degradation of the data and thus causes additional loss of information**, **a re-regularization scheme has been developed which allows resampling without additional loss of in-formation**. For a typical ClONO profile retrieved from spectra as measured by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), the constrained retrieval has 9.7 degrees of freedom. After application of the pro- posed transformation to a coarser information-centered altitude grid, there are exactly 9 degrees of freedom left, and the averaging kernel on the coarse grid is unity. Pure resampling on the information-centered grid without re-regularization would reduce the degrees of freedom to 7.1 (6.7) for a stair-case (triangular) representation scheme." Even, in the case of using both regridding and the re-regularization scheme the "meaningful" height of retrieval cannot and will not exceed the "meaningful" height of the OEM retrieval.

Moreover, it cannot be argued that the "meaningful" retrieval height has increased. The cut-off value of 60% in the uncertainty plot which is proposed by the authors as a response to the first reviewer (comment 6) is an ad-hoc value. This being said, the authors are well-aware of the importance of having a quantitative cut-off height, as in their own words as the response to the second referee about the advantage of the OEM they have written that: "The maximum valid height can be chosen using the averaging kernel values mathematically and does not require the ad hoc removal of the top 10 to 15 km of the profile." yet as is clear the proposed maximum valid height of retrieval has no mathematical validity (unlike the OEM acceptable height). Additionally, the uncertainty **of 60% for lidar measurements is considerably large**, and as has already been pointed out in the previous comments, in Jalali et al, 2018 the authors changed the cut-off height to 0.9 based on the fact that they believed the uncertainty of about 10% is huge. As the authors mention in one of the responses to the second referee the uncertainty of 40% is closer to the OEM cut-off (and still much larger than the OEM uncertainty). I agree with the authors that the smaller uncertainty of the OEM retrieval is due to the **information** from the a priori**, thus taking the a priori out equals taking some information out and should result in lower acceptable height of retrieval and larger uncertainty**. A retrieval with higher uncertainty and lower acceptable height with respect to OEM has no value.

In summery the proposed method has no advantage in comparison with the OEM or traditional method. The effect of a priori is still present and the uncertainty has significantly increased. Although the authors provide several more nights which is appreciated **still they insist on a false claim of gaining higher acceptable altitudes (specially in Fig. 4)**. As seen in Fig.3 of response to the second referee in 50% of the cases for PCL the proposed method does not reach a higher altitude therefore, it is not clear that under what conditions the proposed method results in the so-called higher acceptable retrieval heights. The statement such as "It is up to the researcher to decide if the coarse grid heights are appropriate or not given the characteristics of their data set." provided to the responses to the second referee is not scientific. The authors have mentioned that "The goal of this method is remove the a priori influence from the final retrieval and it is successful" which I fail to see how a method with considerably higher uncertainty, and lower meaningful height of retrieval can be considered a successful method to remove a priori.