Response to Referee 1:

Comment: I thank the authors for their thorough and considerate replies to the reviewers comments. They addressed enough of our concerns for this paper to be suitable for publication after a few minor adjustments. I have a conceptual disagreement with the authors on the finer points of optimal estimation (e.g. I have produced several climatological datasets that contain a priori information and they have been used to evaluate trends; one has to be careful about not overusing the prior but it's possible and, in my opinion, preferable). However, the preference to minimise the influence of a priori information on data is held by the majority of scientists I've met and this isn't the place for that fight. I hope to one day encounter the authors at a conference and exchange opinions over a drink.

Reply: We appreciate this balanced and tolerant stance.

Comment: Minor comments: P5L22 I don’t see a clear statement of A = GK.

Reply: We have added the above equation to the paper as Eq. 6.

Comment: Fig.6 The x-scale of this diagram makes it impossible to judge the similarity between the lines at most heights. plusminus 200 should be fine, with an annotation to indicate the extreme outlier.

Reply: We have changed this as you suggested. The new figure is below and we have replaced it in the paper.

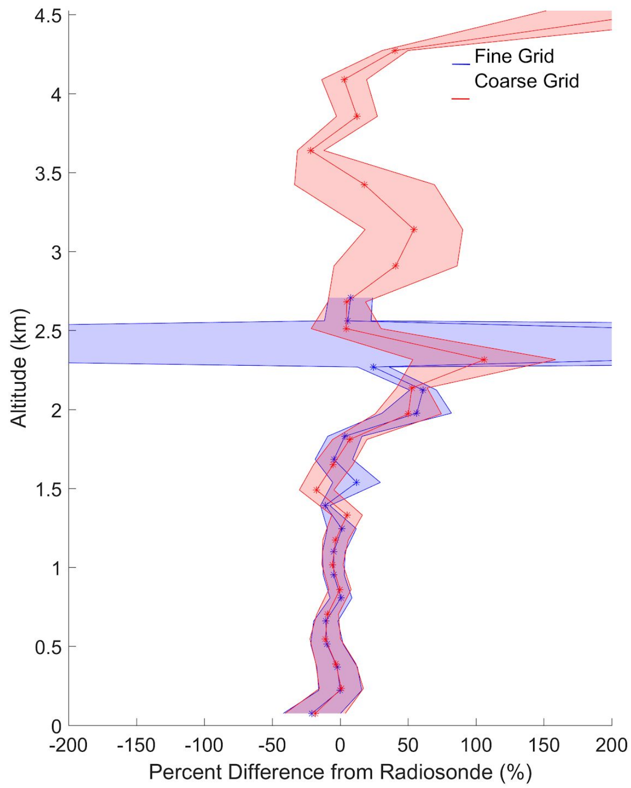

Figure 6: The relative percent difference between the radiosonde and the fine and coarse grid retrievals on 22 January 2013 1200\,UT. The 1-sigma uncertainties for percent difference are shown as shaded regions. The fine grid results are shown in blue and the coarse grid results in red. The largest percent difference for the fine grid is 600\% and is not shown in this figure.

Comment:
Grammatical suggestions:
P2L5 ‘their retrievals, with a much finer grid spacing than passive’
P3L7 ‘some foundational material’
P5L2 Perhaps ‘considered by’ rather than ‘included in’,
(4) The first S is italicised.
P7L1 ‘the fact that the uncertainties of OEM describe a different thing.’ Error and uncertainty are different concepts; you’re talking about the latter.
P7L8 ‘same as the number of retrieval levels.’
P7L16 ‘kernel contains information regarding’
Tab.1 ‘The second column are the elements’
1Fig.2 ‘freedom for the retrieval are 8.2’

P12L15 ‘through the mixing ratio formulae of Hyland and Wexler (1983).’
P15L1 I believe Poisson is capitalised as it refers to a Frenchman.
P24L3 ‘SNR of 2 and 10 km from the top of’
P28L8 Rayleigh should be capitalised.

Thanks for spotting the errors. We have applied your corrections to the paper.

Response to Referee 3:

Comment:

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.

Referee 3 is concerned that we have not provided a complete response to some of their previous comments. We think this misunderstanding arises about use of language in our new method. We will present a summary of our detailed reply to Referee 3's suggestions, but begin with a summary of our disagreements.

First some clarification of how we use terms in our manuscript. Our retrieval method is iterative, using the Levenbert-Marquardt method. In the manuscript "iterations" refer to the number of times the Levenberg-Marquardt algorithm loops to calculate a retrieval relative to specified convergence criteria. In our manuscript, a "run" refers to a complete OEM or ML retrieval, which typically would require 10 iterations each in order to converge. The second run of the OEM using the course grid is called a Maximum Likelihood solution, because we set the covariance of the *a priori* profile to infinity, that is Sa^-1 = 0.

We run the OEM code twice, first on a fine grid then on a coarse grid, using a non-linear Levenberg-Marquardt scheme for both runs. There is no switching of the method inside one retrieval. For example: for the temperature retrieval, in the first run of the OEM on the fine grid, 8 iterations occur to calculate the OEM temperature retrieval. We then calculate the coarse grid from the averaging kernels of the first run using VCG's triangular method to find the appropriate grid points. Then, the second retrieval is run on the coarse grid with Sa^-1=0, and the same Marquardt-Levenberg retrieval algorithm is used iteratively to obtain a solution.

Therefore, the only difference between the results of the two runs are due to the covariance matrix and the grid-width.

We do not switch methods inside one OEM run between iterations of the Levenberg-Marquardt method, thus, the grid on which the Jacobians are evaluated does not change during a retrieval run.

We have not forced the averaging kernels to be 1 in the second run (that is using ML method on the coarse grid retrieval). The averaging kernels are calculated automatically by the retrieval algorithm. We have included the averaging kernels of the second run to verify that the *a priori* information is minimized in the final results, as evidenced by the averaging kernels being of order unity.

We hope this summary clears up any misunderstanding concerning our methodology. A detailed response to Reviewer 3's concerns is provided below.

Comment: 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:
*z*ˆ^ = [(KW)^TS_e^-1 KW]^-1(KW)^TS_e^-1t=G_ML y, (1)
x^ - x = ()x + WG_ML epsilon. (2)
Where z^ is the reduced presentation of the retrieval, WG_ML K−I **I** is the smoothing error, and A = WG_MLK 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.

Reply: A maximum likelihood solution was defined by R. A. Fisher in the 1920s and it is defined as the solution of an inverse problem without formal use of *a priori* information. Fisher explicitly states that this solution does not provide the solution of maximum *a posteriori* probability but as the solution best compatible with the measurement it is a solution in its own right. The retrieval on a coarse grid is the natural way to obtain a maximum likelihood solution. Admittedly, any discrete representation of the solution of a continuous profile is a constraint in itself, but the effect of this constraint is, contrary to the effect of a formal regularization term, obvious and thus does not need any formal evaluation. The averaging kernel of such a maximum likelihood solution on the coarse grid is unity by definition, and it correctly describes that the effect of the (nonexistent) regularization on the truth as represented on this coarse grid. The reviewer seems not to be satisfied with this; he seems to prefer instead the effect of the (nonexistent) regularization on the absolute, infinitely resolved truth to be characterized and thus refers to Rodgers' formalism. This, however, is a bit like praying water and drinking wine. This is because even the smoothing error on a finer grid does not characterize the difference between the fine grid retrieval and the truth, but only the difference between the fine grid retrieval and a representation of the truth on this fine grid (See von Clarmann, ``Smoothing error pitfalls'', Atmos. Meas. Tech., 7, 3023-3034, 2014.). Why should it be allowed to ignore the smoothing error component implied by the discrete sampling on the fine grid but not on the coarser grid? Wouldn't it be more consistent to understand the averaging kernel as the characterization of the retrieval relative to a given grid? We tolerate it if the reviewer prefers another construal of what the averaging kernel is meant to be, but in turn we expect that our construal that all retrieval diagnostics refer to a representation of the truth on a certain grid (as opposed to the absolute truth) is tolerated as well.

Comment: 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).

Reply: You are correct a weakly non-linear approach is required. That is why we use the Levenberg-Marquardt method. We have added a sentence stating this in the methodology section.

Comment: Later when using the nonlinear method, the Jacobian matrix should be transformed as: K'=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.

Reply: As stated in our general summary for the third reviewer, the second retrieval is made on the coarse grid. We do not need this transformation of the Jacobian because the coarse grid Jacobian is directly provided on the coarse grid by the forward model used in the second ML run.

Comment: 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.

Reply: The reviewer seems not to have understood the purpose of our method. The purpose is to present the retrieval on the coarse grid, because then all resolution issues are obvious directly from the grid definition and the problem that the resolution deviates from the sampling is avoided. If we suppose the diagnostics to represent the difference of the retrieval with respect to the truth as represented on the coarse grid and not with respect to the absolute truth (which is impossible anyway) this causes no problem. This representation has several advantages which have been thoroughly discussed in the preceding rebuttals. One of them is, that multiple results represented in our way on the same grid have all the same altitude resolution which allows easy analysis of time series. In the traditional Bayesian representation, the resolution typically changes from profile to profile (via the state-dependence of the K-matrix and in consequence the changing weight between the measurement term and the regularisation term of the G matrix) and such a time series is not physically meaningful because it involves comparison of apples and oranges.

Comment: Finally, the representation error epsilon_r=K(I−WW*)x should be calculated as an additional error.

Reply:

The representation error is a meaningful quantity only if your reference is the truth as sampled on the fine grid (which obviously is not the absolute truth). If we conceive our error analysis as a characterization of the expected difference between the measurement and truth as sampled on the coarse grid, the representation error is meaningless. Of course one can calculate the representation error, but it is kind of ad hoc quantity because it depends on how fine the fine grid is.

The authors' view is to conceive error analysis as a characterization of the result with respect to the truth as sampled on the retrieval grid (thus without representation error) but other views are equally justifiable.

The equation quoted uses the variable x, which is the true profile, which we do not know. We cannot replace the true x by the x retrieved on the fine grid here (because this includes a priori information and noise). The only chance would be to evaluate it in a statistical sense:

S_representation = K(I−WW*) S_a (K(I−WW*)x)^T

which involves a climatological a priori covariance matrix S_a; I see the problem that the climatological variability on the fine grid, specific for the given measurement location, might not be well enough known to calculate a reliable representation error.

Comment: 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).

Reply: The choice of which space is adequate for a comparison depends on which characteristics of the results are to be discussed. It would not be appropriate to use the discussed transformation in this case since the two results are calculated independently on two different grids.

Comment: Furthermore, the general form of the maximum likelihood which is presented in this section is an irrelevant topic to the paper.

Reply: The general form of the maximum likelihood method as we present it in the Maximum Likelihood section is relevant because we use it for our retrieval on the coarse grid. We do not see why it should be irrelevant.

Comment: The general form for the cost function can be written as:
cost=(y-F(x))^T S_epsilon^-1 (y-F(x))+Regularization term (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.

Reply: This statement holds only if we stay in the same grid. Transformation to an information-based coarser grid can make an ill-posed retrieval sufficiently well-posed.

Comment: 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:
x_ML = (S^ ^-1 - Sa^-1)^-1[S^^-1 x_MAP -Sa^-1 -x_a].(4)
However, this solution is only valid if K_l^T S_epsilon^-1 K_l = S^ ^-1 -S_a^-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:
W^T R W = 0 (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_f retrieval. 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).

Reply: First, a transformation of a given retrieval to another grid as presented by von Clarmann and Grabowski is an apt method if the computational effort of re-doing the retrieval on another grid is a serious obstacle, e.g., in the case of large amounts of satellite data. Within linear theory, the von Clarmann and Grabowski resampling/re-regularization procedure and a new retrieval on the coarse grid are equivalent. The re-regularization scheme of von Clarmann and Grabowski corresponds exactly to the coarse grid retrieval basis (grid along with interpolation scheme).

Comment: 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.

Reply: The von Clarmann and Grabowski staircase regularization matrix emulates in the fine grid a coarse grid retrieval where values between two altitude levels remain constant. Conversely their triangular regularization matrix emulates a coarse grid retrieval based on linear interpolation between two altitude levels. Since we do not use the resampling approach but do a direct new retrieval on the coarse grid employing a forward model which internally does linear interpolation of the state variables with altitude, this corresponds to the von Clarmann and Grabowski triangular case.

Comment: 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.

Reply: We strongly disagree. A transformation of a fine grid solution to a coarse grid using the von Clarmann and Grabowski approach and a new retrieval on the coarse grid are equivalent within linear theory. Which of both methods is more adequate depends on practical considerations (Computational effort versus coding effort and risk of coding errors). Both provide the same amount of information, distributed in the same way over altitude, and both lead to a unity averaging kernel on the coarse grid. We use the first step of the von Clarmann and Grabowski method to determine a vertical grid coarse enough to allow a solution without formal constraint. The next step is to find the solution on this coarse grid. This can be done either by a maximum likelihood retrieval on this grid, or by transforming the retrieval in the von Clarmann and Grabowski sense, where the original prior is taken out and replaced by a regularisation which emulates the representation on the coarse grid.

Comment: 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.

Reply: The statement that the maximum meaningful height cannot exceed that of OEM is true only in a Bayesian world where the only meaningful question is "what is the width of the *a posteriori* uncertainty?" We, however, insist that the question "What can we learn from the measurement without involvement of any a priori information?" Is a legitimate approach, too. With this question in mind it is fair to discard Bayesian results which contain a certain fraction of *a priori* information. The related threshold will always be an ad hoc decision and may vary from application to application. A measurement with a certain *a priori* content simply is not measurement information but a mixture of measurement and a priori information, and there are applications where we do not want this. A Bayesian estimate with a certain *a priori* content can be biased towards the a priori while a maximum likelihood estimate will be an unbiased estimate of the smoothed truth, with the smoothing accomplished only by the coarse grid. There are cases conceivable where the Bayesian bias is intolerable while the degradation in terms is acceptable, and thus the acceptable altitude range can be larger in the maximum likelihood case than in the OEM case.

Comment: 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.

Reply: We do use the degrees of freedom to determine the valid altitude coverage for the maximum likelihood retrieval. We use the method of von Clarmann and Grabowski and make each coarse grid layer so wide that it represents one degree of freedom in a sense that the associated partial trace of the fine grid averaging kernel is unity. There is nothing wrong in using the uncertainty of the retrieval as an acceptance criterion.

Comment: 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

Reply: Here the reviewer again compares apples and oranges. The apparently lower uncertainty of OEM is lower only because it does not represent the pure knowledge included in the measurement but it represents the combined knowledge of a priori confronted with measurement information, as we explained in our previous answers. The OEM uncertainty and the maximum likelihood uncertainty answer different questions and it is a misconception to compare these numbers to conclude which method is better.

Comment: (when plenty of reliable climatologies are available).

Reply: We have learned from David Hume (1739) that what was valid and reliable until yesterday is not necessarily valid today.

Comment: I strongly believe that the proposed method with such high uncertainties is not suitable for trend analysis purposes.

Reply: The uncertainties of our method are not larger than those of OEM, they describe something different. If a climatology is, as suggested by the reviewer, used as *a priori*, any trend will be underestimated because the climatology will make the variability of the estimates artificially smaller. OEM results where the K-matrix (and in consequence the averaging kernel) is state-dependent cause time series where each profile has different vertical resolution. Thus no meaningful trend can be calculated because each change of the vertical resolution aliases into the value at a certain altitude. Our suggested maximum likelihood method is not affected by any of these problems.

Comment: 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.

Reply: As we have stated in the summary, this is not what we do. As a second step we run a maximum likelihood retrieval with Sa^-1 = 0 on the coarse grid, with the new K matrix directly evaluated on the coarse grid. Due to the choice of Sa^-1 = 0 in the new retrieval the *a priori* is removed as evidenced by the unity averaging kernels shown in the paper The purpose of the fine grid OEM retrieval is to find out, via the associate averaging kernel, what the optimal coarse grid is.

Comment: Again, no mathematical proof is provided to show that how the effect of the a priori is minimized.

Reply: The proof goes as follows:
A = d ^x_coarse / d x_true,coarse = (K^T Se^-1 K + S_a^-1)^-1 K^TSe^-1K;
with Sa^-1 = 0 we have
A = (K^T Se^-1 K)^-1 K^TSe^-1K = I_coarse
the dependence of the result on the a priori = (I-A)x_a = (I-I)x_a = 0
q.e.d.
We do not claim that this unity averaging kernel is apt to calculate the smoothing error with respect to the absolute truth but only with respect to the truth as represented on the coarse grid. But this holds for any smoothing error: It is never related to the absolute truth but always to the representation of truth on a given finite grid.

Comment: 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.

Reply: This is not true - the answer was provided to Reviewer #1's comment for P19L24 that the temperature cutoff has already been discussed in Jalali et al. 2018 with regards to the Temperature climatology and was found to be adequate. We apologize for not placing it also in the second response to Kgaran.

Comment: 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)

Reply: There is no hard science behind the cutoff used by Sica and Haefele (2016), or for that matter have international programs such as the Network for Detection of Atmospheric Climate Change (NDACC) specified one. The cutoff reflects the confidence one has in the a priori values used. For instance we are currently retrieving relative humidity from lidar measurements using the ERA-5 re-analysis for a priori temperature and water vapour profiles using derived covariances (using radiosondes and ERA-5). In this case, at some times and altitudes, the re-analysis is better than the lidar measurement for temperature, and the response function is less than 0.8, but the retrievals are still good, as the lidar is contributing some information to the retrieval. In the pilot study we did mentioned above demonstrating the method, we used the US Standard Model as the a priori water vapour profile and we had little confidence in it (this model gives a single water vapour profile for all times, locations and heights), hence we cutoff the retrieval at 0.9. In the end what is important is how your retrievals compare with other independent, validated measurements; in this manuscript the agreement between the radiosondes and lidar measurements is excellent.

Comment: "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 heightheight 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.

Reply: What is appropriate can be application-dependent.

Comment: 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 proposed 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."

Reply: We don't do a mere resampling onto the coarse grid but an independent maximum likelihood retrieval on the coarse grid; the latter is emulated in the von Clarmann and Grabowski approach by the combined re-regularization/resampling method.

Comment: 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.

Reply: For applications where any *a priori* content beyond the implicit constraint by the grid is pernicious - and there are a lot - maximum likelihood can provide a larger valid altitude range.

Comment: 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).

Reply: The mathematical validity is by definition: when the a priori is removed it doesn't contribute. What is important is the practical validity. The constrained ML solution on the coarse grid is constrained by the initial OEM to have an a priori contribution much less than the statistical uncertainty of the measurement, unlike in the traditional method where the a priori can contribute 10s of degrees of temperature uncertainty with no ability to determine how bad the contribution may be. Please see Sica and Haefele (2015) where the pressure *a priori* contribution is discussed in detail. For more insight into the traditional method please see the traditional method "best practices" document. (Leblanc, T., et al. (2016). Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms -- Part 3: Temperature uncertainty budget. Atmospheric Measurement Techniques, 9(8), 4079-4101. doi:10.5194/amt-9-4079-2016).

Comment: 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.

Reply: First, the authors mention in the paper that reducing the choice of uncertainty threshold to 40% would decrease the height down to the same height as the fine grid 0.9 cutoff height in both the daytime and nighttime cases.

Second, the random uncertainty for the ML retrieval for water vapour is not the same uncertainty as the OEM uncertainty for temperature. The former is based solely on the measurements, and the second includes the *a priori* and is by definition smaller.

Third, the uncertainty for 60% is large, but was picked as the maximum possible uncertainty that one might want to accept. This choice of the threshold is incumbent upon the researcher to decide and should be determined based on their application of

the method and the quantity they wish to retrieve. Uncertainties for water vapour are not the same as uncertainties for temperature measurements.
The merit of this manuscript is not based on how much higher the retrieval reaches, but the advantage of the method itself and the success of removing the prior contribution. As stated previously, a trend analysis is limited by the prior's contribution and the variable averaging kernels over a data set. This method allows the researcher to process an entire data set using a representative coarse grid with no prior influence and constant vertical resolution.
The authors will add a sentence to the summary reiterating the fact that the 60% threshold is a choice and that 40% would produce results closer to the original cutoff heights. We believe this is fair to the reader and leaves them with the choice to decide what threshold is appropriate for their use.

Comment: 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.

Reply: We strongly disagree. For some applications prior content is harmful while random error bars are not. Any better statistical tool can deal with error bars but there are hardly any which can deal with a priori content. We concede that for some applications the maximum a posteriori solution is the adequate one but why should the question for the pure measurement information be off-limits? The OEM error bar is, rigorously speaking, not the error bar of of the measurement but of the measurement conflated with the a priori. To ask for something else is forbidden only under Bayesian theory.

Comment: A retrieval with higher uncertainty and lower acceptable height with respect to OEM has no value.

Reply: It has value in that it provides the result based on purely measurement information and no *a priori* contribution. Again, for trend analysis this is extremely important since the changing vertical resolution with each retrieval must be accounted for when using OEM. Our method provides the necessary tool for researchers to avoid the problem of *a priori* contribution and variable vertical resolutions across an entire data set. It is useful regardless of whether or not the last valid height increases or decreases. It is up to the researcher to determine what their acceptable threshold is depending on their desired retrieval quantity, the region of the atmosphere in which they are observing, and the design of their instrument.

Comment: In summery the proposed method has no advantage in comparison with the OEM or traditional method.

Reply: We have clearly stated why we disagree with the reviewer.

Comment: The effect of a priori is still present...

Reply: We have sufficiently explained in several of the comments as to why this is not so. The *a priori* is indeed removed as evidenced by the unity averaging kernels shown for both the temperature and water vapour retrievals.

Comment:... and the uncertainty has significantly increased.

Reply: As discussed above, the reviewer comares apples and oranges. What we have done is that we have isolated the pure measurement information. This is only possible on a coarser grid to avoid ill-posedness of the inverse problem. Direct comparison of the OEM and the maximum likelihood uncertainties is misleading.

Comment: 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).

Reply: What is acceptable depends on the acceptance criteria. We have clearly stated what criteria we use and these criteria are adequate for our purpose, namely, to test up to which altitude we can get a measured profile without sizeable a priori contamination. The attempt to isolate measurement information from the Bayesian mixture of measurement and a priori is not an inadequate one. That being so, the wording "false claim" sounds quite hostile.

Comment: 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.

Reply: We will reword this as follows "The decision if the coarse grid heights are appropriate or not given the characteristics of their data set depends on the particular application tha data user has in mind and the question the data are supposed to answer."

Comment: 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,..."

Reply: We have already commented on this above

Comment: and lower meaningful height of retrieval

Reply: We do not agree with the reviewer what is to be called "meaningful".

Comment: can be considered a successful method to remove a priori.

Reply: The averaging kernel is unity on the coarse grid. This indicated that the removal of the formal a priori has been successful.

A practical information-centered technique to remove *a priori* information from lidar OEM retrievals

Ali Jalali[1], Shannon Hicks-Jalali[1], Robert J. Sica[1,2], Alexander Haefele[2,1], and Thomas von Clarmann[3]

[1]Department of Physics and Astronomy, The University of Western Ontario, London, Canada

[2]Federal Office of Meteorology and Climatology, MeteoSwiss, Payerne, Switzerland

[3]Forschungszentrum Karlsruhe, Institut fur Meteorologie und Klimaforschung, Karlsruhe, Germany

Correspondence to: sica@uwo.ca

Abstract. Lidar retrievals of atmospheric temperature and water vapour mixing ratio profiles using the Optimal Estimation Method (OEM) typically use a retrieval grid with a number of points larger than the number of pieces of independent information obtainable from the measurements. Consequently, retrieved geophysical quantities contain some information from their *a priori*, which can affect the results in the higher altitudes of the temperature and water vapour profiles due to decreasing signal-to-noise ratios. The extent of this influence can be estimated using the retrieval's averaging kernels. The removal of formal *a priori* information from the retrieved profiles in the regions of prevailing *a priori* effects is desirable, particularly when these greatest heights are of interest for scientific studies. We demonstrate here that removal of *a priori* information from OEM retrievals is possible by repeating the retrieval on a coarser grid where the retrieval is stable even without the use of formal prior information. The averaging kernels of the fine grid OEM retrieval are used to optimize the coarse retrieval grid. We demonstrate the adequacy of this method for the case of a large power-aperture Rayleigh scatter lidar nighttime temperature retrieval and for a Raman scatter lidar water vapor mixing ratio retrieval during both day and night.

1 Introduction

Rodgers (2000) introduced an Optimal Estimation Method (OEM) based on information theory for use in atmospheric remote sensing retrievals. The OEM has primarily been used in passive remote sensing (Rodgers, 1976; Cunnold et al., 1989; Boersma et al., 2004) and it was not until recently that the OEM was applied to lidar measurements to retrieve atmospheric aerosol properties, temperature, and water vapor profiles (Povey et al., 2014; Sica and Haefele, 2015, 2016). OEM is advantageous for lidar work not only because the desired geophysical quantities are retrieved (e.g. temperature, water vapour mixing ratio, etc.) but also because it produces averaging kernels and a full uncertainty budget on a profile-by-profile basis. The averaging kernel matrix is a diagnostic tool that indicates the degree to which the retrieval is determined by the lidar measurements or by the retrieval *a priori* values.

Lidars have high temporal and spatial resolution compared to passive remote sensing instruments, coupled with high signal-to-noise (SNR) ratio measurements over much of their dynamic range, and thus have averaging kernels close to unity for the majority of their retrievals, with a much finer grid spacing than passive instruments. At most retrieval altitudes, the majority of the information comes from the lidar measurements. However, near the top of the lidar retrieval range, and in other regions where the SNR is low, the *a priori* contribution to the retrieval increases and consequently the amount of information from the measurement decreases. The *a priori* influence at the top of the retrieval should be considered when comparing OEM lidar measurements, particularly if different *a priori* profiles are used.

An estimate of the measurements' contribution to the retrieval, otherwise known as the "measurement response", can be calculated by taking the sum of the averaging kernel functions. The measurement response is calculated by multiplying the averaging kernel matrix, $\mathbf{A}$, with a unit vector, $\mathbf{u}$, which we will refer to henceforth as $\mathbf{Au}$. The *a priori* contribution is then 1 minus the measurement response.

An example of the *a priori*'s influence is shown in Fig. 1 of Jalali et al. (2018). Jalali et al. (2018) used more than 500 nights of measurements from the Purple Crow Lidar in London, Ontario between 1994 and 2013 to calculate the OEM temperature climatology. The cutoff height used for the climatology was the altitude at which the measurement response equaled 0.9, or where the retrieval is roughly comprised of 90% measurements and 10% *a priori* information. In order to see the influence of the *a priori* on the temperature retrieval, temperature profiles from two different models, CIRA 86 and the US Standard Model (Committee on Extension to the Standard Atmosphere, 1976), were chosen to use as *a priori* temperatures. Temperatures were retrieved using both *a priori* profiles, and the differences between the two were compared at the altitudes where $\mathbf{Au} = 0.9$ and $\mathbf{Au} = 0.99$. The distribution of the influence of the *a priori* at these altitudes for the entire climatology is shown in Fig. 1 of this paper. However, the temperature *a priori*'s effect is always one or two degrees smaller than the random uncertainties at these altitudes.

The mean value of the histogram at the altitude where $\mathbf{Au} = 0.99$ is $0.53 \pm 1.29\,\mathrm{K}$ and the mean at $\mathbf{Au} = 0.9$ increases to $0.96 \pm 3.25\,\mathrm{K}$. There is a positive bias in both histograms due to the fact that the monthly CIRA-86 temperature profiles are consistently warmer than the yearly US Standard Model profile. The effect of the *a priori* increases as the values of $\mathbf{Au}$ decrease. Also, all values in the histogram are within two sigma of the statistical uncertainty of the PCL climatology.

As Rodgers (2000) suggested, it is important to pick the most accurate *a priori* for the retrieval. We used the CIRA-86 and US Standard Model to investigate the influence of the choice in *a priori* more clearly, as the differences between these two model temperatures profiles is large. If *a priori* profile values from the CIRA-72 and CIRA-86 models had been chosen for comparison, the mean values on the histogram would have been much smaller.

Several methods for reducing the *a priori*'s influence on the retrieval have been suggested by Vincent et al. (2015), Ceccherini et al. (2009), von Clarmann and Grabowski (2007) and Joiner and Silva (1998). Their method to minimize the effect of the *a priori* was based on transforming a regularized to a maximum likelihood retrieval by moving from a fine grid to a coarser grid. Our work applies the methodology of von Clarmann and Grabowski (2007) (henceforth vCG) to a Rayleigh lidar OEM

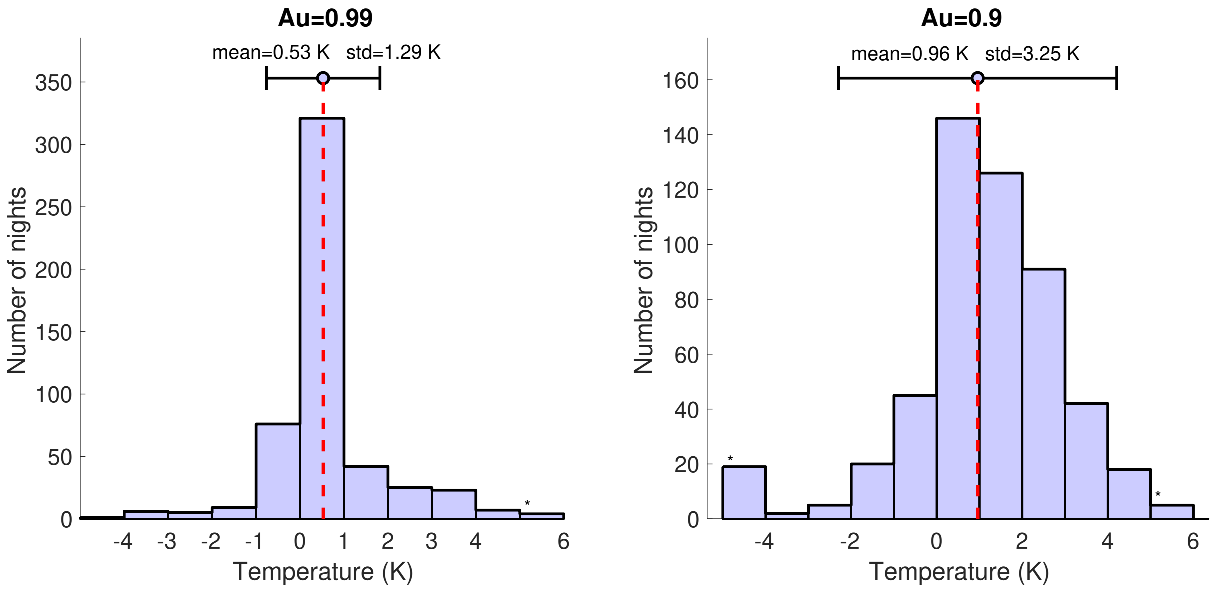

Figure 1. Distribution of the differences in temperatures retrieved at the altitudes where the sum of the averaging kernels ($\mathbf{Au}$) is 0.99(a) and 0.9(b) using two *a priori* temperature profiles - the US Standard Model and CIRA-86 for over 500 nights as detailed in Jalali et al. (2018). The red dashed line shows the mean. For each case, the difference in temperatures is always smaller than the statistical uncertainty at the same altitude.

temperature retrieval and a Raman lidar OEM water vapour retrieval. The method uses a grid transformation on the retrieved temperature and water vapor lidar profiles to remove the *a priori* temperature and water vapour contribution. The transformation is applied in such a way that each final grid point carries roughly one degree of freedom (information-centered). Then, the retrieved profiles are calculated on the coarse grid by re-running the OEM in a way that the effect of the *a priori* constraint is minimized.

We have used two lidars in this study, whose specifications are discussed in more detail in Section 2. Section 3 summarizes some basic foundation fundamental material of the OEM which will be referenced throughout the paper. Section 4 discusses the *a priori* removal methodology with a simple example. The method is then applied in Section 5 for three cases: Raman water vapour daytime, Raman water vapour nighttime, and Rayleigh nightly temperature retrievals. Section 6 discusses the differences between our practical application and the method in vCG and some of the proposed method's advantages. Sections 7 and 8 are the Summary and Conclusions respectively.

2 Description of the lidar systems

Two lidars were used in this study, the RAman Lidar for Meteorological Observation (RALMO) in Payerne, Switzerland and the Purple Crow Lidar in London, Ontario. RALMO was used for the water vapour daytime and nighttime retrievals and the PCL was used for the Rayleigh temperature retrievals.

2.1 RALMO

RALMO is located at the MeteoSwiss research station in Payerne, Switzerland ($46.81°$ N, $6.94°$ E, 491 m a.s.l.). RALMO was built at the École Polytechnique Fédérale de Lausanne (EPFL) and was designed as an operational lidar for model validation and climatological research. RALMO uses a 355 nm wavelength laser operating at 30 Hz with a nominal power of 300 mJ. Measurements are made in one-minute intervals with an altitude resolution of 3.75 m. A typical 30 min water vapour profile will extend to 10 - 12 km at night and 4 - 5 km during the day. Detailed specifications for the RALMO can be found in Dinoev et al. (2013) and Brocard et al. (2013). The water vapour retrieval for daytime and nighttime followed the same procedure as described in Sica and Haefele (2016), with the exception that we now retrieve the overlap, which is no longer a model parameter. Only raw (uncorrected) photocount measurements are used for the water vapour retrievals. The lidar input measurements are 30 min profiles beginning at the same time as the coincident radiosonde launch from the Payerne station. The US Standard Model water vapour profile is used as the water vapour *a priori* input for both daytime and nighttime retrievals.

2.2 Purple Crow Lidar

The Purple Crow Lidar (PCL) is located at the Environmental Science Western Field Station ($43.07°$ N, $81.33°$ W, 275 m a.s.l.) near The University of Western Ontario in London, Canada. The PCL uses a 532 nm wavelength Nd: YAG laser with 1000 mJ per pulse power at 30 Hz. The PCL is comprised of two Rayleigh channels, a High Level Rayleigh (HLR) channel whose high gain detector is useful from between 40 to 110 km and a Low Level Rayleigh (LLR) low gain channel, which is nearly linear due to the use of a neutral density filter, above 25 km. Returns from below 25 km are blocked by a mechanical chopper which controls the firing of the laser. The backscattered photons are collected by a 2.65 m diameter liquid mercury mirror. The temporal and spatial resolution of the PCL is one minute, or 1800 laser shots, and 7.5 m, respectively. The details of the PCL OEM Rayleigh temperature retrieval are discussed in Sica and Haefele (2015) and its application to the PCL data set in Jalali et al. (2018). The PCL OEM temperature profiles are created using nightly integrated HLR and LLR measurements and typically reach up to 100 km. The *a priori* temperature profiles are the CIRA86 (Fleming et al., 1988) and US Standard Model temperatures.

3 Theoretical background

3.1 OEM

The Optimal Estimation Method (OEM) is an inverse method based on Bayesian statistics which calculates the maximum *a posteriori* solution by minimizing a cost function involving both the fit residual and the difference between the result and the *a priori* information. The measured signal $\mathbf{y}$ can be represented as:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \epsilon, \tag{1}$$

where $\mathbf{y}$ is the measurement vector which includes measurement noise (ϵ), $\mathbf{F}$ is the forward model, $\mathbf{x}$ is for the state or retrieval vector and $\mathbf{b}$ is a vector including all model parameters which are included in considered by the forward model but not retrieved. Note that all vectors and matrices will be in bold font, but vectors will be written in lower case and matrices will be capitalized in the same format as Rodgers (2000).

The OEM assumes Gaussian probability density functions (PDFs) to maximize the *a posteriori* probability of the atmospheric state, given the value of the measurements ($P(\mathbf{x}|\mathbf{y})$) and choice of *a priori*:

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}. \tag{2}$$

The possible values of measurements and solutions are distributed by the PDFs $P(\mathbf{y})$ and $P(\mathbf{x})$ respectively, and $P(\mathbf{y}|\mathbf{x})$ is the probability of the measurement given the atmospheric state $\mathbf{x}$. The solution can be optimized in a number of ways depending on the goal of the observer. The method implemented by Sica and Haefele (2015) and Jalali et al. (2018) picks the most likely state for the solution by minimizing a cost function. The cost function (Eq. 3) is a weighted least squares regression with a regularization term comprised from measurements and *a priori* components.

$$Cost = \left[\frac{1}{2}(\mathbf{y}-\mathbf{F}(\mathbf{x},\mathbf{b}))^T{\mathbf{S}_\epsilon}^{-1}(\mathbf{y}-\mathbf{F}(\mathbf{x},\mathbf{b}))\right] + \frac{1}{2}(\mathbf{x}-\mathbf{x_a})^T{\mathbf{S_a}}^{-1}(\mathbf{x}-\mathbf{x_a}), \tag{3}$$

where $\mathbf{x_a}$ is the *a priori* value for the retrieval vector $\mathbf{x}$ and $\mathbf{S_a}$ is the corresponding covariance matrix. The first term in the cost function is the weighted least squares minimization problem, or the fit residuals. Minimizing the cost function produces the retrieval solution ($\hat{\boldsymbol{x}}$), where the solution is then the maximum *a posteriori* solution based on the PDFs and is given by

$$\hat{\mathbf{x}} = \mathbf{x_a} + (\mathbf{K}^T{\mathbf{S}_\epsilon}^{-1}\mathbf{K} + {\mathbf{S_a}}^{-1})^{-1}\mathbf{K}^T{\mathbf{S}_\epsilon}^{-1}(\mathbf{y}-\mathbf{F}(\mathbf{x_a})) = \mathbf{x_a} + \mathbf{G}(\mathbf{y}-\mathbf{F}(\mathbf{x_a})), \tag{4}$$

where $\mathbf{K}$ refers to the Jacobian matrix, $\mathbf{G}$ is the gain matrix, and $\mathbf{S}_\epsilon$ is the covariance matrix of the error measurements. The gain matrix describes the sensitivity of the retrieval to the observations:

$$\mathbf{G} = \frac{\partial\hat{\mathbf{x}}}{\partial\mathbf{y}} = (\mathbf{K}^T{\mathbf{S}_\epsilon}^{-1}\mathbf{K} + {\mathbf{S_a}}^{-1})^{-1}\mathbf{K}^T{\mathbf{S}_\epsilon}^{-1}. \tag{5}$$

One of the advantages of the OEM is that in addition to obtaining a retrieval/solution vector, the method also provides diagnostic tools and a full uncertainty budget. The primary diagnostic tool is the averaging kernel matrix ($\mathbf{A}$) which represents the sensitivity of the retrieved state to the true state (Eq. 6).

$$\mathbf{A} = \mathbf{GK}. \tag{6}$$

At each retrieval grid point (level or altitude), the averaging kernel shows the sensitivity of the retrieval to the measurement. The full width at half-maximum of the averaging kernel at each altitude represents the vertical resolution. Eq. 4 can be rewritten using the averaging kernel as

$$\hat{\mathbf{x}} = \mathbf{x_a} + \mathbf{A}(\mathbf{x} - \mathbf{x_a}) + \mathbf{G}\epsilon. \tag{7}$$

Equation 7 shows that if the $\mathbf{A}$ is the identity matrix the retrieval is sensitive only to the measurements, with no contribution from the *a priori*. Wherever the row-sums of $\mathbf{A}$ (at each level or altitude) is less than unity, the *a priori* is contributing to the retrieval and the extent of its contribution can be estimated using the measurement response. The averaging kernel also provides a means of calculating the number of degrees of freedom (dgf) in the retrieval by evaluating the trace of $\mathbf{A}$,

$$dgf = \mathrm{Tr}(\mathbf{A}). \tag{8}$$

Ideally, the contribution of the *a priori* is zero at all levels, and dgf equals the number of levels of the retrieved water vapour or temperature profiles.

3.2 Maximum likelihood solution

A maximum likelihood (ML) solution is an inverse technique which does not make use of *a priori* information and finds a solution which is solely based on the measurement information. If a Gaussian probability distribution of measurement errors is assumed, the maximum likelihood solution is the solution which minimizes the squared covariance-weighted differences between the measurements and the forward model (Eq. 1):

$$Cost_{ML} = (\mathbf{y} - \mathbf{K}\mathbf{x})^T \mathbf{S}_\epsilon{}^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}). \tag{9}$$

The solution to the ML inverse problem is then:

$$\mathbf{x} = (\mathbf{K}^T \mathbf{S}_\epsilon{}^{-1}\mathbf{K})^{-1}\mathbf{K}^T \mathbf{S}_\epsilon{}^{-1}\mathbf{y}, \tag{10}$$

which is equivalent to the first term of the OEM solution without regularization. From Eq. 10, the gain matrix for ML is:

$$\mathbf{G_{ML}} = (\mathbf{K}^T \mathbf{S}_\epsilon{}^{-1}\mathbf{K})^{-1}\mathbf{K}^T \mathbf{S}_\epsilon{}^{-1}. \tag{11}$$

Therefore, by definition, the averaging kernel of the maximum likelihood solution must be equal to the identity matrix.

We see that it is possible to arrive at the maximum likelihood solution mathematically through the OEM solution by setting $\mathbf{S_a^{-1}} = \mathbf{0}$ in Eq. 4. Additionally, as the solution is based on Gaussian probability distributions, the uncertainties are calculated in the same manner as in the OEM. However, the maximum likelihood uncertainties will be larger than the OEM uncertainties due to the removal of the inverse of the covariance matrix from the gain matrix, as the *a priori* information no longer constrains the covariance of the retrieval to that of the *a priori* profile. This is not a shortcoming of the ML solution but simply reflects the fact that errorthe uncertainties of OEM designate different things. The OEM uncertainty estimate describes our combined *a priori* and measurement knowledge while the ML error bars refer to the pure measurement information.

4 Methodology

Our objective in this study is to find a practical method to remove the *a priori* information from the retrieval vector. We have based our work upon the methodology of vCG, and have developed a quick and straightforward method to remove the *a priori* from the lidar retrieval. vCG proposed removing the effect of the *a priori* by using an information-centered grid approach. Each level of the retrieval on the information-centered grid contains one degree of freedom, and therefore, the number of degrees of freedom of the signal is the same as the number of the retrieval levels. In this condition, the formal *a priori* information can be removed without de-stabilizing the retrieval.

To create an information-centered grid that contains close to one degree of freedom per level requires the averaging kernel of the fine grid retrieval. For a lidar, this is either the raw measurement spacing or a grid found by integrating some number of raw measurements into larger bins. Therefore, the first step is to run the OEM retrieval following the same procedures as in Sica and Haefele (2015) or Sica and Haefele (2016) which use a slightly non-linear forward model and solve the retrieval using the Levenberg-Marquardt method (Rodgers, 2000). This produces a temperature or water vapor retrieval along with their respective averaging kernel matrices and uncertainty budgets on the “fine grid” or first retrieval grid. For RALMO water vapour retrievals, the fine grid altitude resolution is 100 m and 50 m resolution for the daytime and nighttime retrievals respectively, and 1024 m for the PCL Rayleigh temperature retrieval. The fine grid averaging kernel contains the information regarding the degrees of freedom of the retrieval along the diagonal elements of the matrix (see Sect. 3). The cumulative trace of the averaging kernel is the total degrees of freedom of the retrieval (Eq. 8).

To illustrate the method, we will give a simple example with the fine grid levels, diagonal components of the averaging kernel matrix, and the cumulative trace of the averaging kernel, as shown in Table 1.

We then use the triangular representation from vCG to create the information-centered grid using the fine grid averaging kernel. First, the cumulative trace of the averaging kernel matrix is used to determine the amount of information needed for each grid point on the coarse grid using Eq. 12:

$$dgf_c = \frac{dgf}{\mathrm{int}(dgf) - 1} \approx 1, \tag{12}$$

where dgf_c refers to the degrees of freedom per level on the coarse grid, dgf is the cumulative trace of the fine grid averaging kernel matrix (Eq. 8), and $int(dgf)$ is the integer value of dgf (e.g. int(4.8) = 4). The degrees of freedom per grid point is determined by dividing the total degrees of freedom by one less than the integer value of the total. For example, if the total degrees of freedom of the retrieval is 8.2, then the degrees of freedom per grid point is 8.2/(8-1) = 1.1 degrees of freedom

Fine Grid Levels	Diagonal elements of $\mathbf{A}$	Cumulative Trace of $\mathbf{A}$
1	1	1
2	1	2
3	1	3
4	1	4
5	0.9	4.9
6	0.8	5.7
7	0.7	6.4
8	0.6	7.0
9	0.5	7.5
10	0.4	7.9
11	0.2	8.1
12	0.1	8.2

Table 1.

A simple example for demonstrating the averaging kernel matrix's role in finding the coarse grid which resembles the typical structure of a lidar temperature retrieval averaging kernel. The first column is the retrieval level and for lidar OEM retrievals is typically an altitude. The second column is the elements along the diagonal of the averaging kernel matrix $\mathbf{A}$. The third column is the cumulative trace of $\mathbf{A}$, where the last value determines the number of degrees of freedom per grid point for the coarse grid using Eq. 12.

per grid point. In the triangular representation the information is spread over $dgf - 1$ grid points because the first and last points remain the same as those in the fine grid. It is then necessary to interpolate the fine grid to the points where the diagonal elements are equal to the appropriate degrees of freedom to create the coarse grid. As each grid point contains an equal number of degrees of freedom, the grid points are distributed irregularly. The final levels which are used in the coarse grid are shown in Fig. 2. In this case, we now have coarse grid points at 1, 2.2, 3.4, 4.6, 6.1, 8, and 12. As the sensitivity of the averaging kernel decreases, the number of points used in the coarse grid increases.

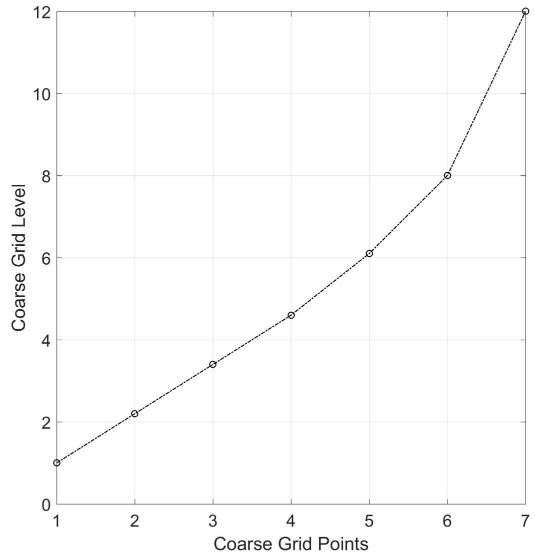

Figure 2. The coarse grid levels are shown for the example case as a function of the cumulative trace of the averaging kernel matrix. The total degrees of freedom for the retrieval is are 8.2, which is spread over the entire retrieval grid such that each point has roughly one degree of freedom. As the SNR of the measurements decreases, more fine grid points are used in the coarse grid, and the distance between points generally increases with altitude.

The resulting coarse grid is then used as the retrieval grid for a second retrieval run. In this manuscript we will refer to a "run" as one retrieval which typically requires 10 "iterations" to converge to a solution. However, before running the retrieval again we remove the regularization term in Eq. 4 by choosing an arbitrarily large *a priori* uncertainty such that the inverse of the *a priori* covariance matrix ($\mathbf{S_a^{-1}}$) becomes zero. If $\mathbf{S_a^{-1}}$ is set to zero, the optimal estimation becomes the unconstrained weighted least squares solution (vCG), which is the solution of the maximum likelihood problem with the assumption of Gaussian residuals in force. The second retrieval is then a ML retrieval which uses the new coarse retrieval grid calculated from the original first OEM retrieval, and the effect of the *a priori* is minimal due to minimizing the regularization term. The ML coarse grid averaging kernels then are unity at all levels.

5 Results

We now apply our information-centered approach, using the triangular representation from vCG, to lidar OEM retrievals in order to minimize the effect of the *a priori*. We will examine the method's effectiveness with RALMO daytime and nighttime water vapor retrievals, as well as with a PCL Rayleigh temperature retrieval. This method is also applicable in general, and can be applied to other lidar retrievals. First, we will discuss the results from the triangular representation and the creation of the "coarse grid" and how it is used as the new retrieval grid. Then we will discuss its effect on the retrieval, vertical resolution, uncertainty budgets, and averaging kernel for a case study for each type of retrieval. We will then discuss the results of the method using representative data sets for all water vapour and temperature retrievals.

5.1 Daytime RALMO water vapor *a priori* removal

5.1.1 Daytime Case Study

The daytime water vapor case study retrieval is a 30 min integration obtained in conjunction with a Vaisala RS92 radiosonde launch from the Payerne station on 22 January 2013 at 1200 UT. This date was chosen because it shows the large impact our method has on low signal-to-noise ratios, which occur during the daytime due to the high solar background or in dry layers (regions with relative humidities less than 25%). The input data grid for this case was binned to 50 m to remove numerical features in the retrieval due to the high background noise levels.

The diagonal values of the daytime case fine grid averaging kernels (Fig. 3a) quickly drop below 1 above 2 km due to a dry layer. The measurement response is shown by the red line which first drops below 0.9 at 2.7 km. This is the uppermost altitude at which we consider the retrieval to not have significant influence from the *a priori*. The coarse grid averaging kernels (Fig. 3b), by design, are all equal to 1 as discussed in Sect. 4 and reach up to 10 km. While the coarse grid ensures that each altitude has 1 degree of freedom, we do not necessarily consider the entire retrieval as meaningful, which will be discussed further below. The vertical resolution of each point on the fine and coarse retrieval grids is shown in Fig. 4. In this case, the fine grid averaging kernels are never exactly 1, therefore have some *a priori* information, which explains why the resolution of the fine grid retrieval is still a little bit coarser than the gridwidth. The vertical resolution of the coarse grid retrieval is still a bit worse. This is attributed to the loss of a fractional degree of freedom, resulting from Eq. 12. The penultimate point in the coarse retrieval grid has a vertical resolution of over 600 m. The coarse grid points which have incorporated more fine grid points have a lower vertical resolution than others (i.e. the points between 2.8 and 10 km altitude).

The daytime water vapor fine and coarse grid retrievals are shown in Fig. 5a and Fig. 5b respectively. The fine and coarse grid retrievals are the same up to 2.5 km, at which point the coarse grid retrieval (in red) begins to more closely follow the path of the radiosonde and the traditional profile (dotted blue) and not the fine grid retrieval (black). The coarse grid retrieval agrees with the radiosonde until 4.5 km. At 4.8 km the statistical uncertainty is above 100%, and the last two points are above 80% statistical uncertainty; therefore, the retrieval is no longer meaningful at these altitudes. All valid points are below the red dotted line. The large peaks in the fine grid retrieval above 5 km show features that are not physical. If we consider the last valid point to be 4.5 km with a statistical uncertainty of 27%, the *a priori* removal method extends the valid altitude range of the daytime OEM retrievals by 2 km.

The three main components of the uncertainty budget are shown in Fig. 5b. The uncertainties shown in this study are relative percent uncertainties, e.g. the uncertainty value divided by the quantity times 100. The fine grid statistical and air density uncertainties increase with altitude due to decreasing SNR of the return photocounts and then decrease as the retrieval falls back to the *a priori* as the signal goes to zero. The coarse grid statistical uncertainties and the uncertainty due to air density continue to increase with altitude, instead of falling back to zero, on the coarse grid because the *a priori* has been removed. The

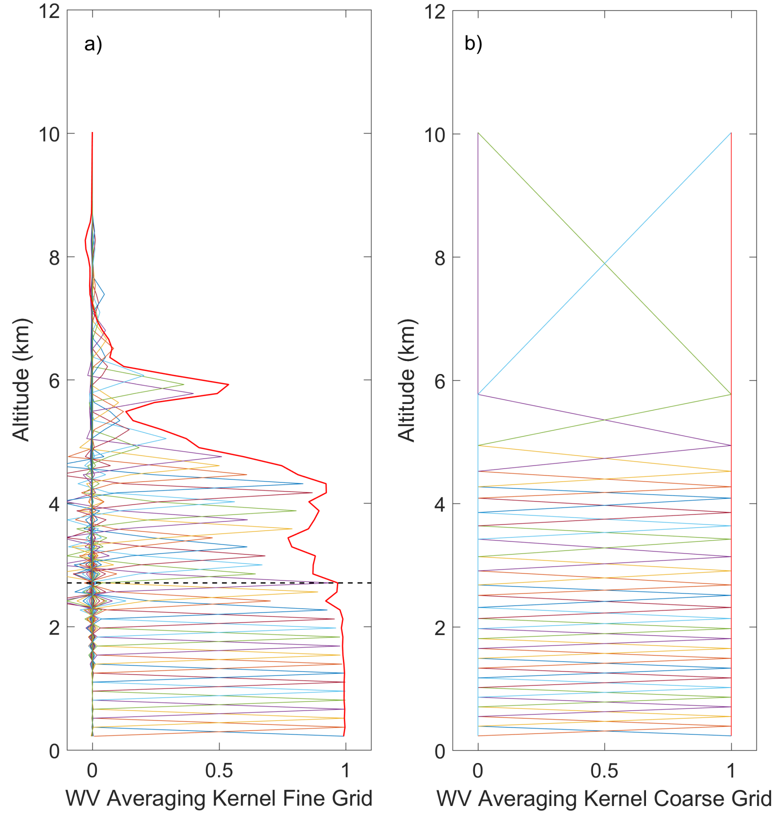

Figure 3. The clear daytime water vapour averaging kernel matrix for 22 January 2013 at 1200UT (a) on the fine grid and (b) on the coarse grid. Every other averaging kernel has been plotted for clarity. a) The measurement response $\mathbf{Au}$ is the red solid line. The horizontal dashed line is the height at which the measurement response is first equal to 0.9 and is the line above which we would consider there to be large influence from the *a priori*. b) The coarse grid averaging kernels all equal 1 and reach up to the last retrieval altitude at 10 km.

a priori has been removed by setting the inverse covariance matrix to zero in Eq. 5. When the *a priori* covariance is removed, the solution space is no longer constrained and the coarse grid uncertainties increase compared to the fine grid uncertainties. The calibration uncertainty also increases, but now remains constant at all altitudes with the exception of the last point, as it is no longer influenced by the *a priori* constraint.

Since the measurement response of the unconstrained coarse grid retrieval is unity everywhere by definition, this quantity is not an adequate criterion for determining the last useful altitude of a retrieval. Therefore, we use the uncertainty of the retrieval as a criterion instead. A relative uncertainty of 60% was chosen as the largest acceptable error, which resulted in a cutoff height of 4.5 km altitude. We found this height to correspond with the altitude at which the signal to noise ratio decreases below 1 and noise begins to dominate the retrieval. However, the choice of the critical uncertainty is a matter of preference, and depending on the goal of the research it may be more preferable to cut the retrieval at a lower uncertainty. It is also important to take the

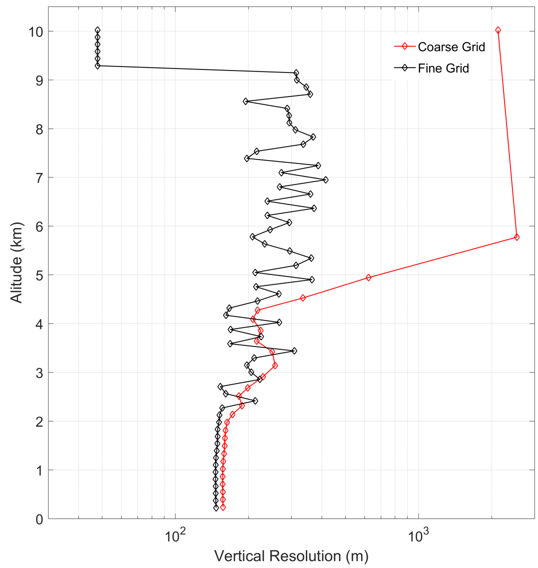

Figure 4. The vertical resolution profile on 22 January 2013 1200 UT. The vertical resolution will decrease on the coarse grid as the points are used to reach one degree of freedom. The last two points have vertical resolutions of several hundred meters, but are not considered meaningful points as they have total uncertainties of larger than 60%.

presence of dry layers into account to avoid cutting the profile too low if the uncertainty threshold is lowered. It may also be more useful to determine a threshold based on absolute errors instead of relative, particularly for the case of dry regions with low signal. To maintain consistency with Sica and Haefele (2015, 2016), we have chosen to use relative errors for this analysis. The second-to-last point in the statistical uncertainty has a mixing ratio uncertainty of 100% due to the lack of signal above 4.5 km. Therefore, the ML coarse grid retrieval was cut to include measurements below 4.5 km. The maximum uncertainty is 46% statistical uncertainty at 3.8 km, where the water vapour signal is very small due to the presence of a dry layer at that altitude. A dry layer is a layer where the water content has below 25% relative humidity. The relative humidity measured by the radiosonde at 3.8 km is 10%. While the *a priori* removal technique increases the maximum retrieval altitude, in addition to removing the contribution from the *a priori* profile, it will increase the statistical uncertainty of the retrieval as well. It should, however, be noted that uncertainties of OEM and maximum likelihood retrievals signify different things. The OEM uncertainties characterize the a posteriori knowledge including *a priori* and measurement information, while the maximum likelihood uncertainties characterize the pure measurement information.

Finally, we compare the fine and coarse grid retrievals with the radiosonde profile in Fig. 6. To highlight the differences in the OEM fine and ML coarse grid retrievals, we have interpolated the radiosonde onto both the fine and coarse grids for comparison and the 1-sigma uncertainties in the percent difference are shown as the shaded regions on each side of the percent difference profile. The radiosonde uncertainties used to calculate the percent difference uncertainties were calculated by propagating pressure, temperature, and relative humidity uncertainties through the 1983 Hyland and Wexler mixing ratio formulae through

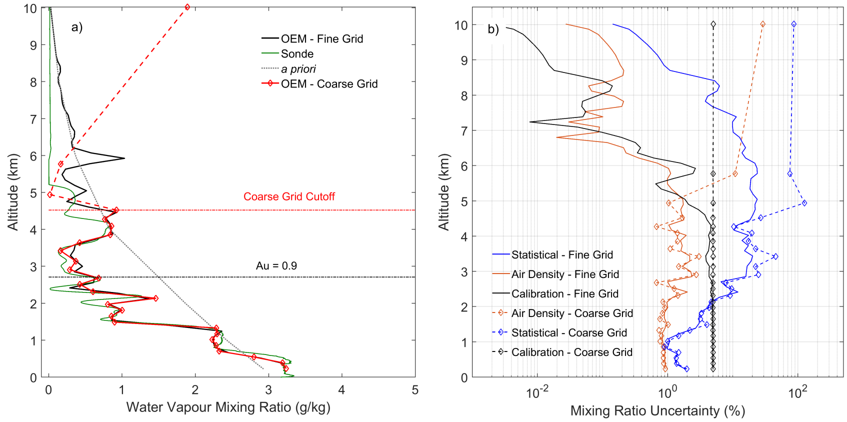

Figure 5. a) The retrieved daytime water vapour profile for 22 January 2013 1200 UT. The fine grid retrieval is in black and includes the *a priori* information. The coarse grid retrieval is in red and the *a priori* (grey) has been removed. The radiosonde is shown in green. The points which we do not consider meaningful because their uncertainties are larger than 80% in the retrieval are shown in dashed red lines. The coarse grid retrieval increases the last valid point by 2 km (red dashed line) and now more closely resembles the radiosonde above the original cutoff altitude of 2.7 km (black dashed line). b) The three primary contributors to the uncertainty budget on January 22 2013 1200 UT are shown for comparison: the statistical uncertainty, the uncertainty due to the calibration constant, and the uncertainty due to air density. The solid lines are the relative uncertainties from the fine grid retrieval, and the dashed lines are from the coarse grid retrieval. The *a priori* begins influencing the profile above 2 km where the uncertainty increases.

the mixing ratio formulae of Hyland and Wexler (1983). The uncertainty values were assumed constant with height using the values presented in Dirksen et al. (2014). The percent difference calculated on the fine grid is cut at the 0.9 measurement response cutoff height. At all altitudes the retrievals agree with the radiosonde within their respective 1-sigma uncertainties. The large uncertainties and the large difference from the radiosonde at 2.2 km is due to the presence of a dry layer where the signal is much weaker. The radiosonde detects much less water vapour compared to both lidar retrievals. That altitude is not included in the coarse grid retrieval due to its lack of information, therefore a similar feature is not seen in the coarse grid percent difference profile.

5.1.2 Daytime Representative Data Set

The *a priori* removal technique was tested on 5 additional days to study the differences between the fine and coarse grid cutoff heights as well as their agreement to the radiosonde (Fig. 7). The daytime water vapour OEM profiles typically reach up to around 3 - 5 km on the fine grid, and up to 6 km on the coarse grid. There is an average of 1.5 km difference between the two cutoff heights. In some cases, the differences are much larger, and this is usually due to the presence of dry layers causing the averaging kernel to decrease at a lower altitude. The large difference between the final altitudes on each grid is typically due to a slow decrease in averaging kernel values with height, as was shown in the case study. Additionally, in some cases, such as on 28 February 2012, the uncertainty never rose above 60%, in which case the second-to-last point on the coarse grid was chosen as the cutoff point.

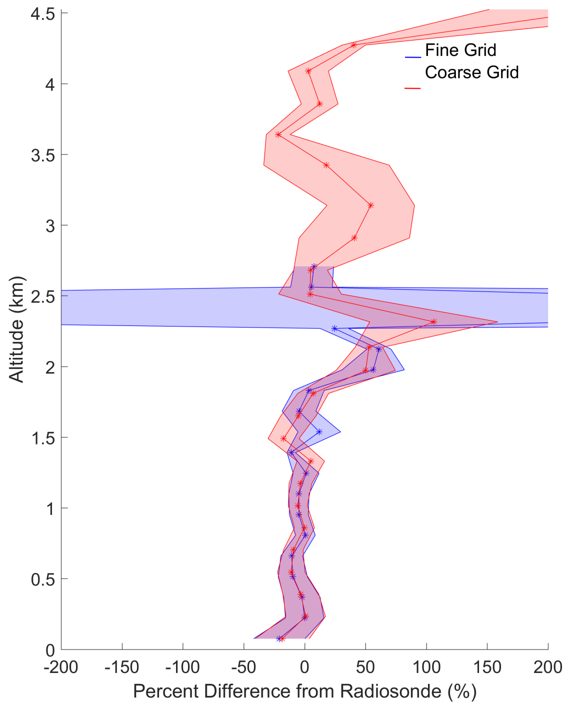

Figure 6. The relative percent difference between the radiosonde and the fine and coarse grid retrievals on 22 January 2013 1200 UT. The 1-sigma uncertainties for percent difference are shown as shaded regions. The fine grid results are shown in blue and the coarse grid results in red.The largest percent difference for the fine grid is 600% and is not shown.

The daytime water vapour OEM fine and coarse grid profiles show similar differences to the radiosonde profile within their respective uncertainties. For each case, with the exception of 5 May 2009, there are very few differences between the fine and coarse grid retrievals from the radiosonde. On 5 May 2009, the coarse grid retrieval was shifted with respect to the fine grid OEM retrieval, possibly due to poor calibration on that day.

The daytime fine and coarse grid retrievals agree with radiosonde measurements within their respective uncertainties and the coarse grid retrievals significantly increase the final meaningful retrieval altitude by an average of 1.5 km. Daytime water vapour retrievals are often limited in altitude due to the high solar background in both the water vapour and nitrogen channels. Increasing the final meaningful altitude by up to 2 km is highly valuable for forecasting and validation purposes.

5.1.3 Examining Cutoff Heights using Signal-to-Noise Ratios

To confirm our choice of cutoff heights for the fine and coarse grid retrievals, we looked at the SNR profiles for the digital water vapour signal for each of the daytime comparisons (Fig. 8). The water vapour signals are roughly 10 times weaker than the nitrogen signal and therefore determine the amount of information available to the retrieval. The SNR profiles were calculated

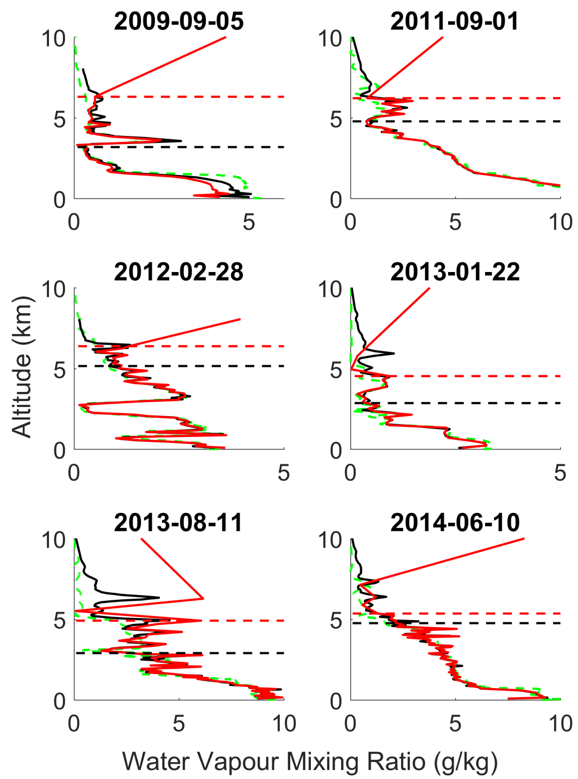

Figure 7. Daytime water vapour mixing ratio retrievals for 5 additional nights. Black lines are the original OEM retrieval on the fine grid, red solid lines are the ML coarse grid retrievals and the dashed green lines are the radiosonde mixing ratio measurements. The black dashed line is the original 0.9 measurement response cutoff height and the red dashed lines are the coarse grid cutoff heights which were chosen as the last altitude whose measurements had less than 60% total uncertainty.

using the raw digital input signals to the OEM retrieval. As digital signals follow Poisson statistics, the SNR was calculated using the following equation:

$$SNR(z) = \frac{N(z) - B}{\sqrt{N(z)}}, \tag{13}$$

where z is altitude, N is the number of photon counts, B is the mean background signal calculated as an average of the counts from 55 - 60 km for the water vapour measurements.

It stands to reason, that as the SNRs of the measurements drop, the OEM dependence on the measurements should also decrease (and the *a priori*'s increase) due to the increase in noise. Typically, the SNR level drops below between 3 and 4 km altitude for daytime measurements due to the high solar background. The 0.9 measurement response cutoff height used for the fine grid OEM results is shown by the blue dashed line in Fig. 8. For each daytime retrieval, the 0.9 measurement response cutoff falls between an SNR of 1 and 2. The green dashed lines are where the measurement response is last larger than 0.8, or

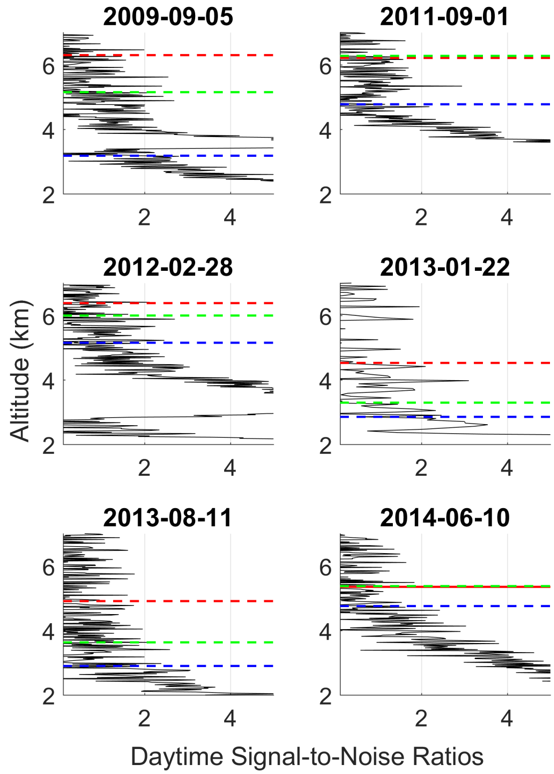

Figure 8. Daytime water vapour SNRs (black). The various cutoff heights are shown in dashed lines. The 0.9 measurement response cutoff is blue, 0.8 measurement response cutoff is green, and the coarse grid cutoff is in red.

the 0.8 measurement response cutoff height. The 0.8 cutoff is consistently located at the heights were the signal-to-noise ratio is unity, and usually 500 m to 1 km or higher than the 0.9 cutoff. The coarse grid cutoff height, shown by the red dashed line corresponds typically to the boundary where the SNR drops below 1 into the region where noise dominates. The location of the coarse grid cutoff then makes sense, as this would be the altitude where no more information could be gathered and the uncertainties increase beyond what we would consider meaningful or useful. The coarse grid cutoff sometimes coincides with the location of the 0.8 cutoff, but is typically below the coarse grid point. The SNRs of the 0.9 measurement response cutoff correspond to the traditional limits of water vapour measurements for the RALMO lidar, which are typically cut where the water vapour SNR drops below 2. Therefore, for a fine grid OEM retrieval, we find that the 0.9 cutoff is a consistent choice with regards to the traditional method. The 0.8 cutoff height could be used, but we would caution against it as it may induce unwanted amounts of a priori water vapour information into the retrieval. The coarse grid utilizes the amount of information available from the measurements to produce an information-centered profile, therefore, we also find its height appropriate as it borders where the noise begins to dominate the measurements.

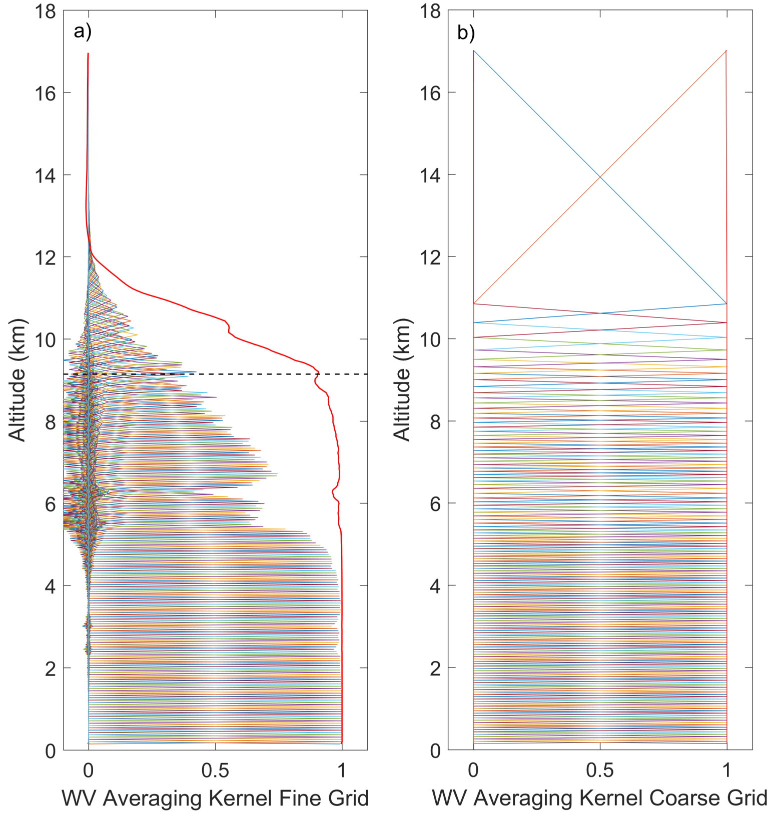

Figure 9. The averaging kernel matrix for the nighttime water vapour retrieval on 24 April 2013 0000 UT. a) The fine grid retrieval with a maximum altitude of 9.1 km (black dashed line). The measurement response is shown in red. b) The coarse grid retrieval, where each averaging kernel is 1 for all altitudes.

5.2 Nighttime RALMO water vapour *a priori* removal

5.2.1 Nighttime Case Study

The nighttime case study retrieval uses a 30-minute integration on 24 April 2013 0000 UT which coincides with the time of radiosonde launch. The fine retrieval grid for the RALMO water vapor retrieval is 50 m.

The averaging kernel matrix for the fine and coarse grid retrievals is shown in Fig. 9a and Fig. 9b, respectively. The altitude where $\mathbf{Au}$ first equals 0.9 for the fine grid retrieval is at 9.1 km, which is typical for a 30 min nighttime measurement. The coarse grid averaging kernels all equal 1, with the second-to-last altitude at 11 km.

Unlike the daytime case, the nighttime vertical resolution between the fine and coarse grid retrievals is very close up to 5 km where they begin to diverge (Fig. 10). This is because the nighttime averaging kernels are very close to 1 until 5 km.

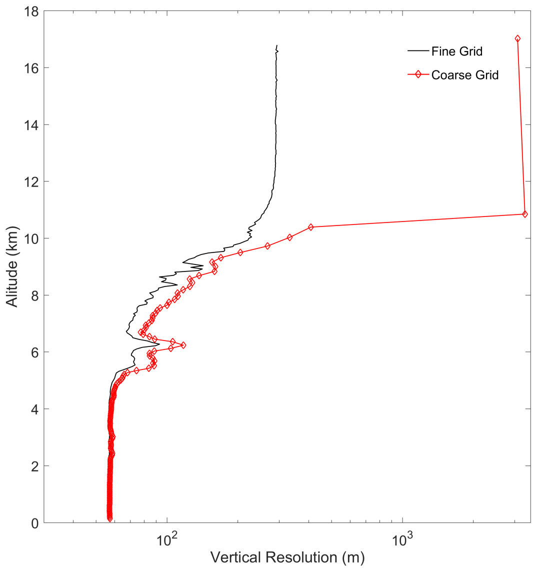

Figure 10. The vertical resolution for April 24 2013 0000 UT. The vertical resolution on the coarse grid retrieval decreases as more points are added to ensure that each bin has one degree of freedom. The coarse grid resolution is shown in red and each point is marked. The fine grid has points every 50 m therefore they are not shown individually.

As the *a priori* enters the signal, more points from the fine grid are used to create the coarse grid, resulting in larger coarse grid averaging kernels and decreasing the vertical resolution. Figure 11 shows the final water vapor retrievals on the fine and coarse grid as well as a Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN) Vaisala RS92 radiosonde profile. Both fine and coarse grid profiles agree past the 0.9 cutoff and up to 9 km at which point the coarse grid retrieval diverges from both the fine grid retrieval and the radiosonde. We do see small differences in dry layers where the signal level is lower, however, the differences are inside the total uncertainty. The last four points in the retrieval are shown in dashed lines because we do not consider them to be meaningful points as their total uncertainties are 70% or larger.

The uncertainties for the nighttime retrievals are shown in Fig. 11b. Similarly to the daytime retrievals, we have shown the top three uncertainty contributors for comparison. Below 5 km the uncertainties are the same, as there is no influence from the *a priori*. However, above 5 km the uncertainties begin to increase due to the removal. The statistical uncertainty increases to almost 100% uncertainty at the second-to-last point due to the lack of signal above 11 km. The mixing ratio uncertainty due to the calibration uncertainty is now constant with altitude, which we would intuitively expect and contributes roughly 5% uncertainty to the mixing ratio measurements. The uncertainty due to air density increases by a maximum of 0.2% at the second-to-last point. We would consider anything above 9.7 km to be invalid since points above that height have a total uncertainty of 60% or higher. The last valid point has a total uncertainty of 52% at 9.7 km. Therefore, the *a priori* removal technique increases the maximum valid altitude of the retrieval by 600 m.

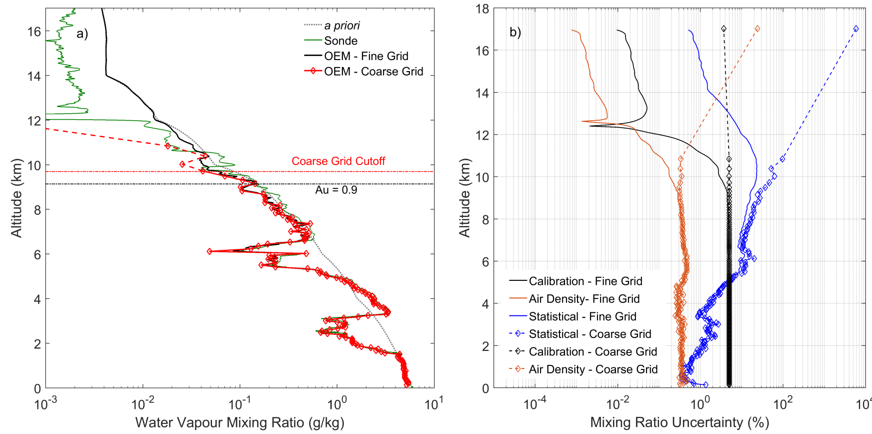

Figure 11. a) The water vapour retrieval for 24 April 2013 0000 UT. The fine grid retrieval is in black, the coarse grid retrieval is in red. In general, both OEM retrievals on the coarse and fine grid, and the radiosonde agree until the original cutoff altitude at 9.1 km (dashed black line). The dashed red lines above 9.7 km show the points we do not consider meaningful due to their large uncertainties. Therefore, the *a priori* removal technique increases the last altitude bin by 600 m. The method is limited by the lack of water vapour in the upper troposphere which causes a large and rapid drop in signal. b) The three largest relative uncertainty components are compared here on the fine and coarse grid. The drawback of the *a priori* removal technique is that while you gain in altitude, you increase the uncertainty. At 9.7 km the statistical uncertainty is 52%, above which is where we no longer consider the rest of the retrieval to be viable.

The fine and coarse grid retrievals do not change very much with respect to each other until 9.1 km where the averaging kernels begin to drop off significantly. They both produce similar differences with the radiosonde (Fig. 12), except between 5 – 7 km, likely due to the dry layer present at those altitudes and smoothing from the coarser grid. The uncertainties for the nighttime percent differences are more variable than the daytime percent difference uncertainties due to the fact that we used a GRUAN RS92 radiosonde on this night which calculates the uncertainties of the radiosonde as a function of altitude. Mixing ratio uncertainties were calculated in the same way as the daytime radiosonde mixing ratio uncertainties.

5.2.2 Nighttime Representative Data Set

The *a priori* removal method was applied to 8 additional nighttime retrievals (Fig. 13). The nighttime cutoff heights in Fig. 13 show a general increase in cutoff height when using the a priori removal method, albeit not as large. As with the daytime retrievals, the coarse grid cutoffs were chosen to be the last altitude below with a total uncertainty less than 60%. Choosing a maximum uncertainty of 40% would result in cutoff heights closer to the original fine grid's. In all cases, the coarse grid increases the maximum acceptable altitude, however, in some cases by only a few hundred meters. On those nights, the averaging kernels decrease quickly after the original fine grid cutoff height, therefore there is very little information with which to create the coarse grid.

In all cases, the water vapour nighttime OEM fine grid and ML coarse grid retrievals produced profiles which agreed with the radiosondes within their respective uncertainties. Differences larger than .4 g/kg, between both retrievals and the radiosonde

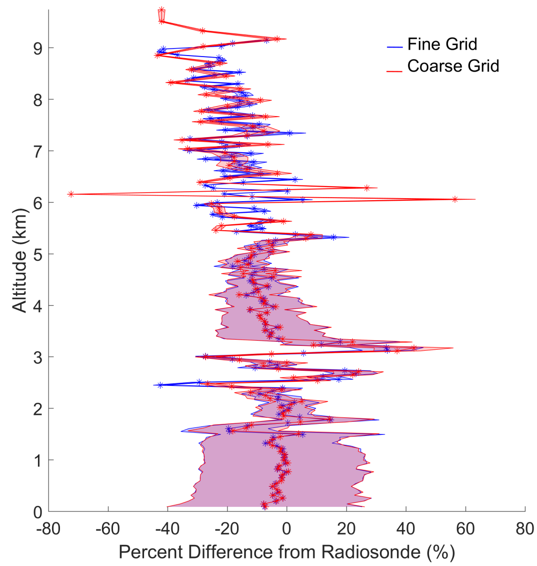

Figure 12. The percent difference from the radiosonde for both the fine and coarse grid retrievals. Both show similar differences with the radiosonde and the last valid height is 9.7 km.

profile, can be seen on 25 May 2012. This was likely due to lack of co-location with the lidar, as the balloon was 10 km away from the lidar at that altitude.

Using the *a priori* removal technique for nighttime retrievals may be helpful when trying to improve water vapour measurements of the Upper Troposphere and Lower Stratosphere (UTLS) region. However, in this case, because the nighttime measurements have large SNRs and a rapid change from high to low signal values, we do not see as large of a difference between the coarse and fine grid retrievals as we do in the daytime retrievals. For nighttime retrievals, the coarse grid may not provide an operational advantage, but can still be used to homogenize a data set for trend analysis or climatological studies which would require no *a priori* influence. This will be discussed further in Sect. 6.

5.2.3 Nighttime Cutoff Heights and SNRs

Similarly to the daytime water vapour measurements, we have also compared the SNR values with the fine grid and coarse grid cutoff heights (Fig. 14). As before, the fine grid 0.9 measurement response cutoff corresponds to the last point where the measurement response is greater than 0.9 and is shown by the blue dashed line in Figure 2. We have also included the 0.8 measurement response cutoff height (green dashed line) for comparison which is calculated in the same way as the 0.9 measurement response cutoff. Lastly, we have included the cutoff height for the coarse grid, chosen as the last height at which the total uncertainty of the retrieval is less than 60%.

In all cases, the 0.9 measurement response cutoff corresponds to a SNR of 2. When we compare the 0.8 measurement response cutoff height with the 0.9 cutoff height, we see that the 0.8 cutoff is typically between a few hundred meters to 1

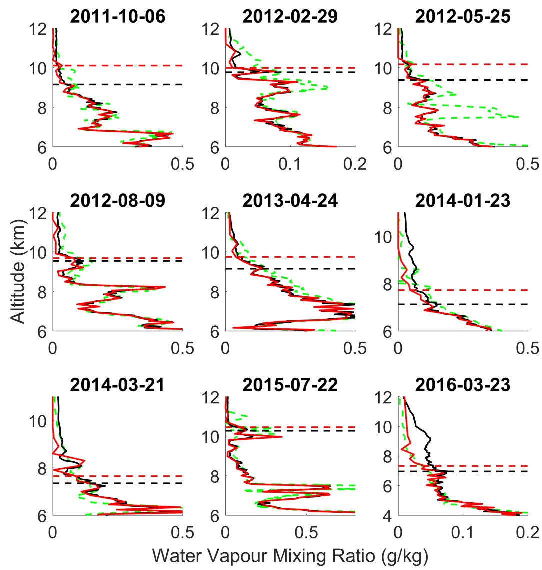

Figure 13. All nighttime water vapour retrievals. The radiosonde is shown by the green dashes, the fine grid retrieval in black, and the coarse grid retrieval in red. The 0.9 cutoff height for the fine grid is shown by the black dashed line while the coarse grid cutoff height is the horizontal red dashed line.

km higher. However, unlike the daytime measurements, the 0.8 cutoff and the coarse grid cutoff are very close and are either close to 1 or at the boundary where the SNR starts to be noise-dominated. Therefore, we would suggest when using fine grid nighttime OEM water vapour retrievals, to use the 0.9 measurement response as a cutoff height since the 0.8 cutoff height may be in the region where noise dominates, which would lead to larger amounts of the *a priori* entering the retrieval.

5.3 Purple Crow Lidar Rayleigh temperature *a priori* removal

We picked a sample night, 12 May 2012, from the Rayleigh temperature climatology in Jalali et al. (2018) to illustrate the *a priori* removal procedure for a Rayleigh temperature retrieval. The original OEM retrieval fine grid was 1024 m, and the *a priori* temperatures were taken from the CIRA-86 model. The details regarding the OEM retrieval are discussed in Sica and Haefele (2015) and its results applied to the climatology are discussed in Jalali et al. (2018).

The averaging kernels for the fine grid and coarse grid retrievals are shown in Fig. 15a and Fig. 15b. The red line is the measurement response or the estimate of the averaging kernel's sensitivity to the measurements. The height at which the measurement response equals 0.9 was chosen as a “cutoff" height in Jalali et al. (2018), which is shown in Fig. 15a with a dashed line. After applying the *a priori* removal, the averaging kernel on the coarse grid is equal to 1 at each point. Fig. 15b shows that at the coarse grid points, according to the averaging kernel, the temperature retrieval is completely sensitive to the measurements and therefore there is no *a priori* contribution.

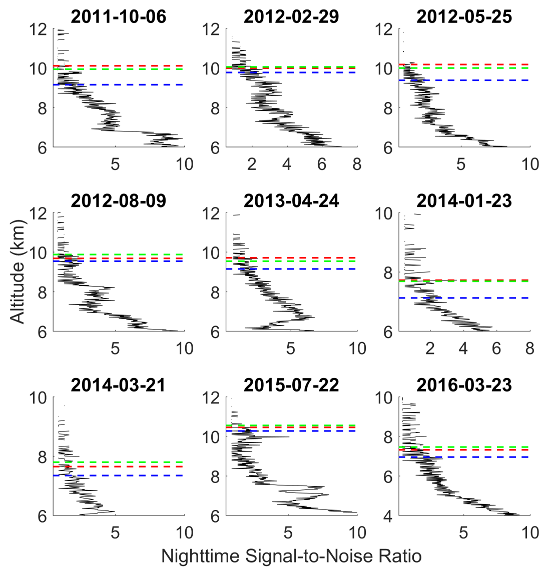

Figure 14. Nighttime SNR calculations for each nighttime water vapour OEM retrieval. The dashed lines are the corresponding cutoff heights: 0.9 measurement response (blue), 0.8 measurement response (green), coarse grid (red).

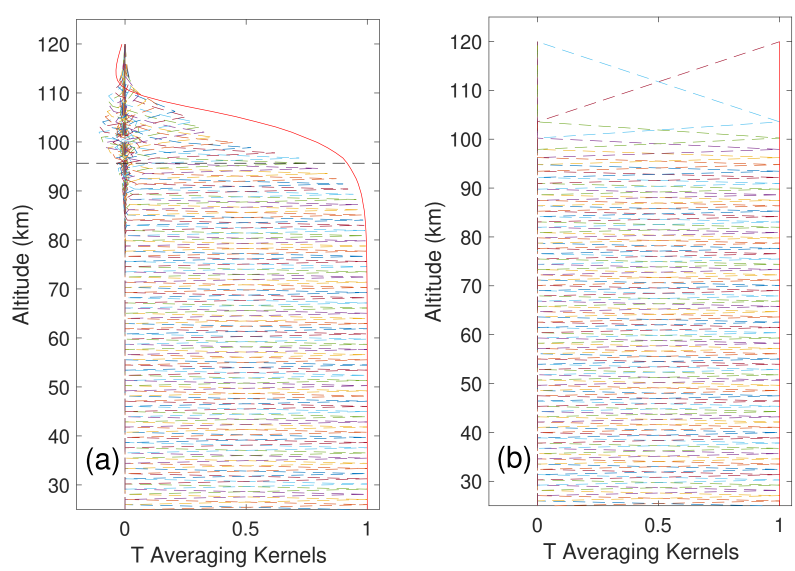

Figure 15. The PCL averaging kernels for the temperature retrieval on 12 May 2012 on the fine grid (a) and on the coarse grid (b). The $\mathbf{Au} = 0.9$ cutoff height on the fine grid is shown by the black horizontal dashed line at 97 km. The red lines on the edges of the averaging kernels are the measurement response. The coarse grid extends the temperature upwards by 4 km.

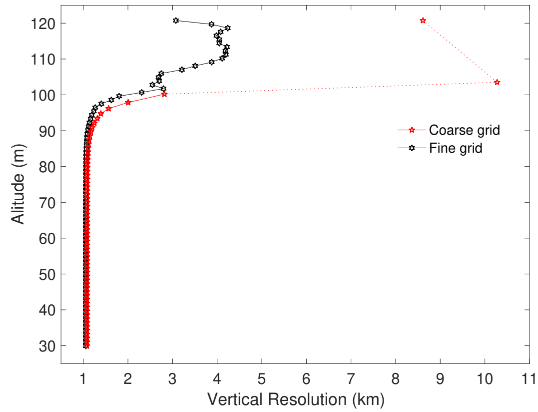

Figure 16. The PCL vertical resolution for 12 May 2012 on the fine and coarse grid. The vertical resolution is similar up to 85 km on both grids. Above this height the vertical resolution decreases until it is 10 km in resolution above 100 km altitude (dotted red line). We consider 100 km to be the highest meaningful point on the coarse grid due to large uncertainties above that height.

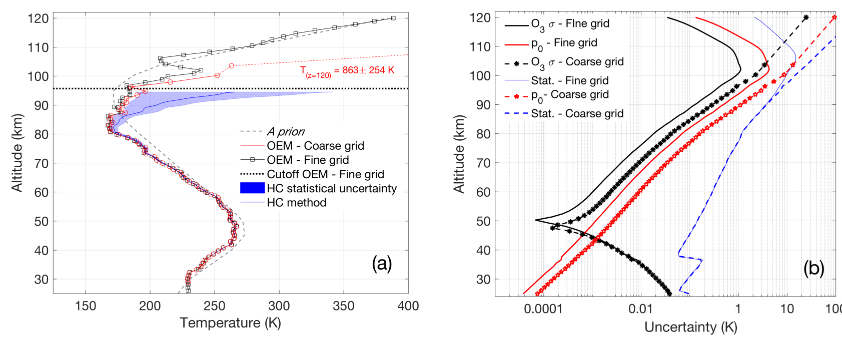

Figure 17. (a) PCL temperature retrieval for the fine and coarse grids on 12 May 2012. The temperature and its uncertainty for the last coarse grid point has a large value and it is not shown. (b) The statistical and systematic uncertainties due to the tie-on pressure and ozone cross section for the PCL temperature retrieval. The other systematic uncertainty terms included in our retrieval are not shown.

The vertical resolution for both grids is similar up to 85 km altitude (Fig. 16). Above this height the coarse grid incorporates more points from the fine grid, and thus, the vertical resolution decreases. The values of the vertical resolution (Fig. 16) of the two highest points for the coarse grid are 10 km at 100 km and 8 km at 110 km. However, the corresponding total uncertainties at these altitudes are above 100% and 60%, therefore we do not consider them to contribute to the retrieval.

Figure 17a shows the OEM fine and ML coarse grid temperature retrievals compared to the Chanin and Hauchecorne (HC) temperature calculation Hauchecorne and Chanin (1980). The two OEM and ML retrievals are identical up to 88 km. Above 88 km the coarse grid retrieval differs from the fine grid retrieval and provides only four additional levels. The last 2 levels are shown with dashed lines in Fig. 17a and are points that we would not consider in the retrieval due to their large uncertainties. The last meaningful point shown in Figure 17a is around 100 km, where the corresponding statistical uncertainty and systematic uncertainties due to the tie-on pressure and ozone cross section are 15, 9 and 2.3 K, respectively (Fig. 17b). Therefore, the last

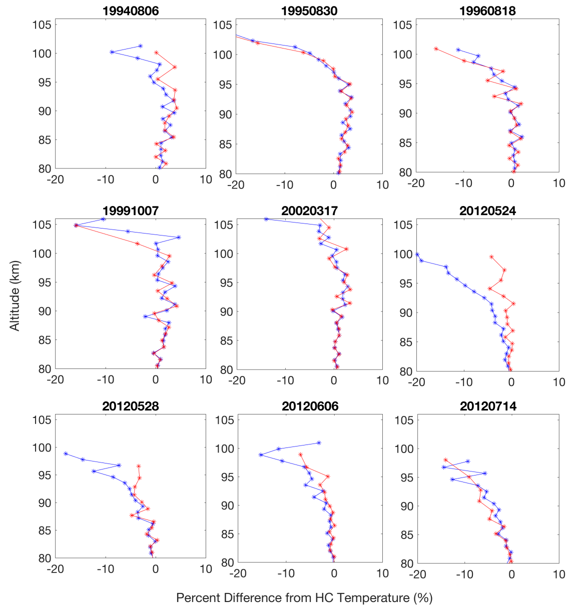

Figure 18. The percent difference between the fine grid retrieval with the HC method (blue line) and coarse grid (*a priori*–removed) with the HC method (red line). Below 80 km the retrievals are identical, as the coarse and fine grid are identical.

valid point of the retrieved temperature on the fine grid is within the total uncertainty of the coarse grid and the final retrieval altitude increases by 4 km.

In this case, it cannot be concluded if the HC result is closer to the fine or coarse grid result. In order to investigate, we used 9 additional nights randomly picked from PCL measurements and the percent difference between the fine and coarse grid retrieval with the HC method was calculated (Fig. 18). In general, the method does just as well as the regular OEM, or better, with respect to the HC method results. We may also conclude that, in general, the a priori temperatures do not have a large effect on the profiles retrieved with the OEM for most nights, however, for nights such as 24 May 2012 and 28 May 2012 the *a priori* seems to have had a larger effect which is removed by our technique.

A consequence of applying this method is that the uncertainties in the retrieval increase where the coarse grid is not equal to the fine grid. Figure 17b shows the statistical uncertainty on the fine and coarse grid, as well as two of the largest systematic uncertainties, including the uncertainty in the retrieved temperature due to the tie-on pressure and ozone cross section. The most sensitive uncertainty parameter is the statistical uncertainty, which changes from 13 K to 20 K at 98 km. The details of the systematic uncertainties on the fine grid are discussed in Sica and Haefele (2015) and Jalali et al. (2018). The systematic uncertainties increase after *a priori* removal due to the gain matrix (Eq. 5) increasing after the regularization term is removed. In general, all uncertainties on the coarse grid (Fig. 17b) increase at higher altitudes, where contribution from the *a priori* starts. The increasing of the random uncertainties at the highest altitudes is due to decreasing photocounts from the exponential decrease in air density.

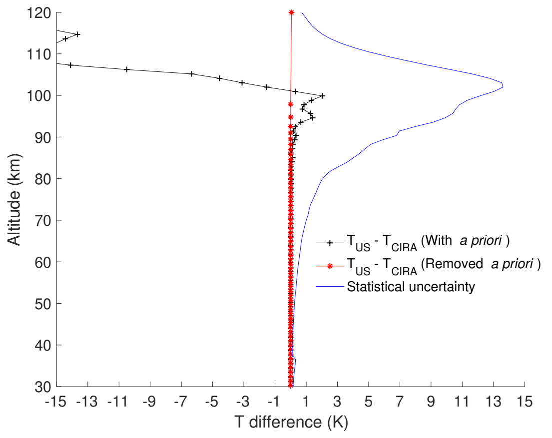

Figure 19. PCL temperature difference between the OEM retrieved temperature profiles using values from the US Standard Atmosphere and CIRA-86 as the *a priori*.

To illustrate that the *a priori* is in fact being removed, we compared the temperature retrievals using two very different *a priori* temperature profiles, one calculated by CIRA-86 and one calculated by the US Standard Model (Fig. 19). The difference between the two temperatures on the fine grid retrieval is shown by the black curve and is about 2 K at the 0.9 cutoff line, within the statistical uncertainty. The difference increases rapidly above that height. The same temperature difference after the *a priori* is removed is shown in red and is on the order of zero at all altitudes.

The HC method considers the fact that the atmosphere consists of isothermal layers and uses a seed pressure (or temperature) at the top of each measurement profile to calculate the temperature in the lower layers. The maximum height that there is enough information in the signal is at SNR equals 2. Therefore, the seed value usually is chosen at the altitude that SNR of 2 and top 10 km from the top of the temperature profile is removed due to the seed value uncertainty. We also examined the relationship with the Rayleigh temperature retrieval and the SNR of the Rayleigh channel signal to determine if there was a similarly consistent value associated with the measurement response cutoff height as there was for the water vapour retrievals. However, based on the examination of all 500+ nights in Jalali et al. (2018) study, removing 10 km below the altitude at which the SNR=2 yields cutoff altitudes higher than the measurement response of 0.8, which suggests that removing 15 km instead of 10 km may be more consistent with the OEM technique.

6 Discussion

We have developed a method to remove the influence of the *a priori* temperature and water vapour profiles on the retrieval based on the method discussed in vCG. These authors presented a method to re-regularize the retrieval in a way that the original *a priori* information is removed and the regularization on the fine grid emulates a coarser grid. These reregularized profiles can then be resampled on a coarse grid without additional loss of information. The optimal coarse grid is determined from the averaging kernel matrix of the original retrieval. This method effectively removes the prior information from the retrieval while keeping the retrieval stable by the use of the coarser final grid. This independence of *a priori* information can be diagnosed by the averaging kernel matrix, which is unity on the coarse grid.

vCG presented two approaches, a “staircase" representation, and a “triangular" representation, to transform the retrieval from the fine to the coarse grid. The cumulative trace of $\mathbf{A}$ shows the total degrees of freedom of the retrieval. In these representations, the cumulative trace of the averaging kernel matrix $\mathbf{A}$ as a function of altitude is calculated and is then interpolated to the coarse grid based on the centered information approach. As each space contains only one degree of freedom, the spaces are distributed. The staircase representation with its discontinuities at the layer boundaries is not a realistic representation of the atmosphere; therefore we use the triangular representation here to create the coarse grid. In the triangular representation, the highest and lowest level of the coarse grid are considered to be the same as the fine grid and the rest of the grid points are distributed such that each layer between two levels represents approximately one degree of freedom.

Our method differs from vCG in that we do not re-regularize the retrieval to remove the *a priori*. Instead, after the initial retrieval, we remove the regularization term from the retrieval and re-run the retrieval using the coarse grid. This second run of the retrieval is then equivalent to a Maximum Likelihood retrieval whose results are solely based on the information provided by the measurements. Both the proposed method and that of vCG are equally effective; however, our method is more of a brute-force technique but easier to practically implement since it is trivial to re-run the retrieval a second time.

For lidars, the triangular coarse grid calculation results in a grid that is very close to the original OEM retrieval at the lower retrieval altitudes where there is more signal and the averaging kernels of the OEM are close to unity. However, at higher altitudes, where the OEM averaging kernels decrease, the information is spread over more altitudes and therefore the coarse grid spacing becomes larger to compensate for the lack of information. An information-centered re-gridding approach is important for an ML retrieval because it is not guaranteed that any inhomogeneous grid will produce a stable *a priori*-free retrieval. Additionally, a statistical gridding approach is easily automated and creates a grid that represents the physical conditions of the atmosphere.

We have shown how the *a priori* removal method works for three sample retrievals: water vapor during both daytime and nighttime, and a nighttime Rayleigh temperature. The *a priori* removal technique is most useful when the SNR is low, such as for daytime water vapour measurements. The method can increase the daytime retrieval altitude by up to 2 km which is highly beneficial for meteorological studies that rely on accurate tropospheric measurements. The nighttime water vapor retrieval was provided for contrast to illustrate how the *a priori* removal technique does not provide significantly more information when the signal level falls off rapidly.

For Rayleigh temperature retrievals, we used measurements from the PCL in London, Ontario. Jalali et al. (2018) suggested that the 0.9 level be used as the valid cut-off height. In the case of the PCL, we see that the second-to-last point on the coarse grid has a vertical resolution not much larger than the fine grid retrieval (Fig. 16) and is very close to the same height; therefore, the 0.9 measurement response value seems to be a conservative choice for a valid cutoff. We also showed that the effect of the *a priori* is removed completely in the Rayleigh temperature retrieval when we compared the differences in the retrieved temperature using the values from CIRA-86 and from the US Standard Model as the *a priori* profiles (Fig. 19). The presented method provides us with higher altitudes for the retrieved temperature profiles. Additionally, where the retrieved temperature profile in the coarse grid is the same as it is for the fine grid, we can be confident the temperature retrieval has a negligible contribution from the chosen *a priori* temperature profile.

An advantage of our method over OEM is that the entire coarse grid profile is *a priori*–free, in the sense that the regularization term does not contribute to the retrieval. In regions where the SNR is low or the averaging kernel is significantly less than 1, the *a priori* removal method improves the validity of the retrieval. An *a priori*–free profile is especially useful for trend analyses and climatological studies which must not include prior information and must be wholly based on measurements. The advantage of an information-centered grid for a typical measurement may be used for multiple retrievals. A grid which is optimal for one atmospheric state will in most cases be close to optimal for a similar atmospheric state. With this consistent grid choice, the altitude resolution of a multi-year time series will be consistent which is important when working with data over long time periods or conducting trend analyses. Varying information content of the individual measurements will lead to error bars of different size. The coarse grid allows time series analysis or trend analysis for single altitudes without problems caused by varying vertical resolution.

The important trade-off with this technique is that the uncertainties of the retrieval increase when moving from an OEM fine grid retrieval to a ML coarse grid retrieval. Both the systematic and statistical uncertainties in the second ML retrieval increase due to the removal of the inverse of the *a priori* covariance matrix from the gain equation (Eq. 5). The vertical resolution of the profile also increases as a consequence of the method. We also lose the ability to determine the maximum useful retrieval altitude by using the averaging kernels. In this case, it is necessary to use the uncertainties to determine the maximum altitude. While the *a priori* removal gives us more confidence in the retrieval, we may not consider the entire profile meaningful due to high uncertainties. Hence, the last few points with unity averaging kernel value on the coarse grid may not be recognized as valid retrieval levels.

7 Summary

We have developed a practical and robust method which removes the effect of *a priori* information in lidar OEM retrievals. The method utilizes an information-centered coarse grid which is derived using the averaging kernels from the initial "fine grid" retrieval. The resulting coarse grid is then used, alongside setting the inverse of the *a priori* covariance matrix to zero, to create the final ML retrieval without any *a priori* information. The method has little computational cost; the OEM retrieval is extremely fast even on a laptop computer, so having to do the retrieval twice for each profile is not critical. We illustrated the method using a simple example in Sect. 4 and demonstrated the removal method using the water vapour signal from the RALMO and the Rayleigh temperature signal from the PCL. We summarize the results from both of these examples as follows:

1. Figure 1b) shows that 90% of the nights in the temperature climatology from Jalali et al. (2018) had less than a 5 K influence from the *a priori* temperature profiles at the $\mathbf{Au} = 0.9$ cutoff height. Additionally, in all cases the *a priori* temperature influence was less than the statistical uncertainty, as was illustrated in Fig. 6 in Jalali et al. (2018). Although

small, the *a priori* temperature profile does contribute to the retrieved temperature in regions where the measurement response is smaller than 1.

2. The *a priori* removal technique increased the maximum altitude of the water vapour daytime retrieval by an average of 1 km and up to a maximum of 2 km, however, the maximum altitude is on the same order of the fine grid retrieval height if a lower uncertainty threshold is adopted. Both OEM fine grid and ML coarse grid retrievals produced similar differences with respect to the radiosonde which agreed within their respective uncertainties (Fig. 6). While the nighttime coarse grid retrievals did not show a significant increase in cutoff height, they did increase on average by a few hundred meters. The nighttime water vapour averaging kernels decrease quickly with height and therefore have very little information to add to the retrieval thereby resulting in very small increases in altitude when using the coarse grid.

3. Applying the method to the PCL temperature retrieval showed useful retrievals above the $\mathbf{Au} = 0.9$ cutoff height by 2 km, validating the choice of $\mathbf{Au} = 0.9$ for a cutoff made in Jalali et al. (2018) to form their climatology up to an altitude where tie-on pressure effects were minimal. The temperatures below the cutoff height were the same.

4. In all cases, the vertical resolution of the OEM retrieval decreases after *a priori* removal.

5. The systematic uncertainties after *a priori* removal increase roughly by a factor of 2, but remain on the same order of magnitude as before the *a priori* removal. The values of the systematic uncertainties also remain significantly smaller than the statistical uncertainties.

6. The temperature difference between the PCL retrieved temperature profiles using two different *a priori* profiles were used to show the effectiveness of the *a priori* removal method. The temperature difference before removal around the 0.9 cutoff height was more than 2 K, however, this value was zero for the entire range after *a priori* removal.

7. The water vapour measurement response values of 0.9 consistently corresponded to a SNR of 2 for the nighttime retrievals, and between 1 and 2 for the daytime retrievals. Therefore, it is our recommendation that traditional water vapour retrievals be cut at an SNR of 2 to compare with the OEM water vapour retrievals. Additionally, measurement response values of 0.8 or higher corresponded to SNR values of 1 or less than 1, therefore we would not suggest cutting the water vapour retrievals at heights above which the measurement response is less than 0.9.

8. The Rayleigh temperature measurement response 0.9 cutoff height was also compared to the SNR of the Rayleigh signal. However, no correlation could be found between the cutoff height and the SNR value. In fact, removing 10 km below a SNR of 2 tended to correspond to measurement response values of less than 0.8 which suggests that it may be more appropriate to remove 15 km from the altitude at which the SNR = 2 to achieve results more consistent with the OEM.

8 Conclusions

When designing an OEM retrieval, it is often desirable to understand the effect of the chosen *a priori* parameters or profiles. This effect has been explored in detail for satellite-based and passive ground-based instruments, but not for the new area of applying OEM to active-sensing measurements such as lidar. Lidars are high resolution instruments with significant amounts of information available from their measurements, as evidenced by the retrieval averaging kernels. The OEM helps to illustrate the robustness of the lidar data products with the advantage of providing diagnostic tools, such as the averaging kernel and a full uncertainty budget.

The *a priori* removal technique may be helpful for checking the *a priori*'s influence on the retrieval and in determining the appropriate *a priori*. It is also important to note that the differences between the fine grid OEM retrieval and the coarse grid ML retrieval may be smaller if one uses an *a priori* closer to the true atmospheric state. Often, reanalysis model profiles are used as *a priori* for OEM retrievals because they are closer to the atmospheric true state than a climatological profile. However, the nature of the *a priori* profile should depend on the design of the instrument and the goal of the work.

In this study, the US standard model water vapour profile was chosen as the *a priori* profile to accommodate the operational nature of RALMO lidar water vapour measurements which requires a minimal number of dependencies in the code as possible and preferably no need for internet. The CIRA temperature profile was used for the temperature *a priori* because there are very few model temperature *a priori* profiles above 80 km for the PCL and coincident satellite measurements are not always available. Additionally, when conducting trend analyses or climatological studies it may be more useful to use a consistent *a priori* profile throughout the analysis to avoid inducing trends or biases into the results.

The removal method is most operationally useful for lidar measurements with low signal to noise and a slow transition from regions of high signal to low signal. The method is less effective at increasing the maximum retrieval altitude when signal strength changes rapidly, such as when the nighttime water vapour measurements quickly enter the dry upper troposphere or lower stratosphere. However, the method is most useful for homogenizing large data sets for trend analyses. One representative coarse grid would be applied to an entire data set and a ML retrieval would be run remove *a priori* information from all measurements, thereby making them suitable for trends.

In the future, this method will be applied to the entire 10 years of RALMO measurements to retrieve the water vapour day time and nighttime measurements and create a water vapour climatology. We anticipate that this technique will increase the altitude of the daytime water vapour retrievals by several kilometers. It is also our hope that this method may provide statistically significant measurements in the UTLS region. Finally, the RALMO water vapour climatology will be used to find trends.

Author contributions. Ali Jalali was responsible for developing the *a priori* removal method and code as well as manuscript preparation, including the following sections: Introduction, Theory, Methodology, Rayleigh temperature retrievals, Summary, Discussion, and Conclusion. He also applied the method to the Rayleigh Temperature analysis. This work is the second component of his doctoral thesis. Shannon Hicks-Jalali applied the removal method to the water vapour daytime and nighttime analyses as well as helped with manuscript preparation,

including the following sections: Introduction, Theory, Methodology, water vapour daytime and nighttime retrievals, Summary, Discussion, and Conclusion. This will serve as a component of her doctoral thesis. Robert J. Sica was responsible for supervising the doctoral theses and contributed to manuscript preparation. Alexander Haefele was also responsible for supervising the doctoral thesis of Shannon Hicks-Jalali and contributed to manuscript preparation and scientific discussions. Thomas von Clarmann significantly contributed to the scientific discussions which resulted in this paper, helped develop the method based on his original work, and contributed to manuscript preparation.

Acknowledgements. This project has been funded in part by Discovery Grants and a CREATE Training Program in Arctic Atmospheric Science (PI K. Strong) from the National Science and Engineering Research Council of Canada and awards from the Canadian Foundation for Climate and Atmospheric Science. We would like to thank the Federal Office of Meteorology and Climatology, MeteoSwiss, for its support of this project and providing the water vapour lidar measurements.

References

Boersma, K. F., Eskes, H. J., and Brinksma, E. J.: Error analysis for tropospheric NO_2 retrieval from space, J. Geophys. Res., 109, doi:10.1029/2003JD003962, 2004.

Brocard, E., Philipona, R., Haefele, A., Romanens, G., Mueller, A., Ruffieux, D., Simeonov, V., and Calpini, B.: Raman Lidar for Meteorological Observations , RALMO – Part 2 : Validation of water vapor measurements, Atmos. Meas. Tech., 6, 1347–1358, doi:10.5194/amt-6-1347-2013, 2013.

Ceccherini, S., Raspollini, P., and Carli, B.: Optimal use of the information provided by indirect measurements of atmospheric vertical profiles, Opt. Express, 17, 4944–4958, doi:10.1364/OE.17.004944, 2009.

Committee on Extension to the Standard Atmosphere: U.S. standard atmosphere, US Government Printing Office, pp. 1–227, doi:NASA-TM-X-74335, NOAA-S/T-76-1562, 1976.

Cunnold, D. M., Chu, W., R.A.Barnes, McCormick, M. P., and Veiga, R. E.: Validation of SAGE II ozone measurements, J. Geophys. Res., 94, 8447–8460, doi:10.1029/JD094iD06p08447, 1989.

Dinoev, T., Simeonov, V., Arshinov, Y., Bobrovnikov, S., Ristori, P., Calpini, B., Parlange, M., and Van Den Bergh, H.: Raman Lidar for Meteorological Observations , RALMO – Part 1 : Instrument description, Atmos. Meas. Tech., 6, 1329–1346, doi:10.5194/amt-6-1329-2013, 2013.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmospheric Measurement Techniques, 7, 4463–4490, doi:10.5194/amt-7-4463-2014, 2014.

Fleming, E. L., Chandra, S., Shoeberl, M. R., and Barnett, J. J.: Monthly Mean Global Climatology of Temperature, Wind, Geopotential Height and Pressure for 0-120 km, NASA Tech. Memo., NASA TM100697, 85 pp, 1988.

Hauchecorne, A. and Chanin, M.: Density and temperature profiles obtained by lidar between 35 and 70 km , Geophys. Res. Lett., 7, 565–568, doi:10.1029/GL007i008p00565, 1980.

Hyland, R. and Wexler, A.: Formulations for the thermodynamic properties of the saturated phases of H_2O from 173.15 K to 473.15 K, ASHRAE Tran., 89, 500–519, 1983.

Jalali, A., Sica, R. J., and Haefele, A.: Improvements to a long-term Rayleigh-scatter lidar temperature climatology by using an optimal estimation method, Atmospheric Measurement Techniques, 11, 6043–6058, doi:10.5194/amt-11-6043-2018, https://www.atmos-meas-tech.net/11/6043/2018/, 2018.

Joiner, J. and Silva, A. D.: Efficient methods to assimilate remotely sensed data based on information content, Q. J. Roy. Meteor. Soc., 124, 1669–1694, doi:10.1002/qj.49712454915, 1998.

Povey, A. C., Grainger, R. G., Peters, D. M., and Agnew, J. L.: Retrieval of aerosol backscatter, extinction, and lidar ratio from Raman lidar with optimal estimation, Atmos. Meas. Tech., 7, 757–776, doi:10.5194/amt-7-757-2014, 2014.

Rodgers, C. D.: Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, Rev. Geophys., 14, 609–624, doi:10.1029/RG014i004p00609, 1976.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, vol. 2, World Scientific, Hackensack, NJ, USA, doi:10.1142/3171, 2000.

Sica, R. J. and Haefele, A.: Retrieval of temperature from a multiple-channel Rayleigh-scatter lidar using an optimal estimation method, Appl. Optics, 54, 1872–1889, doi:10.1364/AO.54.001872, 2015.

Sica, R. J. and Haefele, A.: Retrieval of water vapor mixing ratio from a multiple channel Raman-scatter lidar using an optimal estimation method, Appl. Optics, 55, 763–777, doi:10.1364/AO.55.000763, 2016.

Vincent, R. A., Dudhia, A., and Ventress, L. J.: Vertical level selection for temperature and trace gas profile retrievals using IASI, Atmos. Meas. Tech., 8, 2359–2369, doi:10.5194/amt-8-2359-2015, 2015.

von Clarmann, T. and Grabowski, U.: Elimination of hidden a priori information from remotely sensed profile data, Atmos. Chem. Phys., 7, 397–408, doi:10.5194/acp-7-397-2007, 2007.