Reply to the referees

Ole Martin Christensen and Patrick Eriksson

First of all, we want to thank the reviewers. Their comments are relevant, concise and helpful. A revised version of the manuscript, that addresses the issues raised by the reviewers, will be uploaded together with this reply. Additionally, a version of the manuscript where changes are color-coded is uploaded to help editor and reviewers to see where changes are made (T_inversions_AMTD_v3_colorcoded).

This reply responds to comments by both reviewers according to the chronology of the text, and motivates the changes made to the manuscript. Referee comments are given in italic.

Title:

Referee #1: *"Title: I am not quite sure about the grammar of the term 'inversion', i.e., what is inverted and what is the result of the inversion. It might be worthewhile to verify with the help of a native English speaker that the term 'time series inversion of spectra' is unambiguous. "*

The title is grammatically fine according to a native English speaker. In the case of what is inverted and what is the result of the inversion, both the measured spectra and the inverted profiles are time-series.

Introduction:

Referee #1: Introduction: *The paper is primarily about a retrieval strategy, and I understand that it is not very relevant that the case study chosen is based on microwave measurements.*
Thus, I find it misleading that the introduction starts with a paragraph on ground-based microwave spectrometers. I suggest to move the first paragraph of the intro after l4 on page 1558, and to reword the intro such that it does not depend on the microwave technique. This will, of course, imply some wording adjustments but the reader is led more directly to the focus of the paper.

We agree with Referee #1 and have rewritten the introduction to better reflect the contents of the paper.

Retrieval methodology:

Referee #1: Eq 1: *No Bayesian nor MAP context has been introduced so far, so the use of the term 'a priori' is somewhat out of context. I suggest to replace xa by x0 . For the moment, this can be any starting value, since no concept involving the term 'a priori' has been introduced yet. Just say wheresoever near Eq 2 that you replace x0 by xa .*

We agree with Referee #1 and have changed the equation and introduced xa right after equation 2.

Referee #1: p1559 bottom: *I am not sure if 'measurement response' is the commonly used term.*
I have also read 'measurement sensitivity' in this context. It is suggested to check with the literature and verify that you use the established term here (The contents is correct, this comment is only about the wording).

Measurement response is a common term for microwave spectrometers, see for example the reference in the paper. However for readability we changed the sentence to "The sum of a row in the AVK matrix is a measure of the retrieval's sensitivity to changes in the state vector and is called measurement response (Baron et al., 2002), or measurement sensitivity. "

Referee #2: 1560- line 14: *It would be good to clarify here that a diagonal Se for this system implies both that there is no channel-to-channel correlation (instrumental baseline) and (I would think) no long-term bias. I don't think that it would be a good idea to include a bias, but, formally, wouldn't that be appropriate?*

We are a bit uncertain what Referee #2 means with long term bias of the spectra, but the referee is correct that a more complex Se matrix can be used, and as such we have decided to add some comments about this in the discussion section. In addition we have clarified that the noise is "uncorrelated between the channels" when we specify the noise covariance matrix.

Referee #1: p1562 top: *Here you present a method to construct the covariances of the a priori state vector, which contains at least some ad hoc elements. Thus the method is no longer a maximum a posteriori method as stated in the paper. I know that I am somewhat pedantic here, but the problem can be avoided by stating "We use the formalism of the MAP method" rather than "We use the MAP method" throughout the paper.*

We agree with the Referee #1 and added to the following text to the manuscript: "It should be noted that since information about the temporal correlation of water vapour at these altitudes is limited, these covariance matrices are created in a somewhat ``ad-hoc" fashion. As such, the covariance matrices should be viewed as illustrative examples rather than a perfect representation of the atmospheric variability.

Referee #2: 1562- line 13-24: *Why is there a correlation length associated with the error in the mean? If something is truly an "error in the mean", doesn't that imply that (in the mean) the error is true always, not just for a long temporal period? E.g., if a particular instrument always measures 5% too low I don't see that this has any temporal correlation length. Perhaps the authors are trying to make a point here about seasonal variations (where the a priori might come from a climatology), but I'm not sure.*

With uncertainty in mean, it is meant the uncertainty of the mean value at that point, i.e. how this value deviates from the true mean in both space and time. To clarify this, the sentence is changed to " The latter uncertainty arises from a limited knowledge of the atmospheric mean state at that particular altitude and time." This also means that a 100% correlation over the retrieval time is not necessarily realistic or correct, an thus we use a long, but not infinitely long correlation length.

Referee #1: p1563 last par: *It is not intuitive that the a priori variance is necessarily smaller for a larger time interval. The last two paragraphs in Rodgers 10.3.1.1. seem to me to refer to a condition where all variation around the mean is random. For a larger interval, the sample mean will of course be closer to the population mean. But what if you have, e.g., periodic (e.g. diurnal) changes: A 'morning' climatology will have smaller variability than an 'all day' climatology. I see that in your case study you don't look down into the troposphere, so diurnal variation of H_2O plays no role here, but since the concept presented in this paper has more general applicability, I suggest to add a caveat on this issue somewhere. Some people might wish apply the method to NO_2 or N_2O_5 ...*

We simply state that the variance of a 48h mean is difference from that of a 3h mean. And if we assume that our description of the variable given by the apriori uncertainty matrix is correct the variability will change as specified in Rodgers 10.3.1.1. However, as the referee states, exactly how to specify the apriori statistics and will off course vary depending on which atmospheric variables that are retrieved, but we feel it is beyond the scope of this paper to discuss how exactly to specify this for atmospheric variables not discussed directly in the paper.

Referee #1: p1565: *Have the simulated spectra been superimposed with artificial noise, or are these noise-free test retrievals? Is the noise only in the S matrix or is S chosen consistent with the noise-superimposed synthetic spectra? The authors are encouraged to be more specific here.*

p1564l20 says: "Note that no thermal noise is added to the actual spectra, but only used to specify the covariance matrix ". So it is already stated in the paper that no actual noise is added to the spectra, but we accept the referees feedback that might not be clear enough, and as such the clarifier *"i.e. the spectra themselves are noise free "* is added to the line.

Referee #1: p1565: *"The single spectrum inversions have no error": Is this because no noise has been added to the synthetic spectra, or is it because the results are constrained so heavily towards the a priori that noise cannot push the results away from the prior?*

This is because no noise is added to the spectra, and that the apriori is equal to the true atmosphere, the sentence has been changed to reflect this comment.

Referee #1: p1565 l21 *Not sure what "unit change" is.*

"unit change" is changed to "sudden doubling of water vapour".

Referee #1: p1566 l12/13: *This is confusing? How is the time domain included in the AVK for single spectra retrievals. I thought that single spectra retrievals are n × n, not nN × nN ? I agree that there should be no information crosstalk in time in single spectra retrievals but I doubt that the AVK is the right diagnostic to show this? Or aren't these real single spectra inversions but formal time series inversions but with no constraint in the time domain, thus behaving like single spectra retrievals. Please be more specific here.*

This is already specified in the method section where we write: "For comparison, inversions are also performed with zero correlation in time (blue curve) to mimic single spectrum (1D) retrievals. Though these retrievals could be done on each spectrum separately, it is chosen, for comparison purposes, to perform the retrievals simultaneously using the same formalism as the time series inversions. This is achieved by using a block diagonal a~priori covariance matrix in the retrievals. ". However, a sentence is added to underline this point: "This is due to the fact that the single spectrum retrievals are performed as 2d retrievals with no correlation in time in the S_a matrix".

Referee #2: 1566- line 20: *Assuming by averaging spectra together the sensitivity is increased (as shown in Figure 2), the reduction in the AVK should not be quite 1/16.*

Text is changed to "...roughly 1/16".

Referee #1: p1568: *Have you thought about the Purser and Huang (grid width divided by diagonal of averaging kernel, c.f. Rodgers Sect.3.4.4) criterion for the resolution? It seems more meaningful to me particular in cases when the 'measurement' response is low but the FWHM still small. Let A be a diagonal values with 0.5 along the diagonal, for illustration what my point is.*

We have not thought abut using the Purser and Huang (P&H) criterion for the solution, so first we want to thank for an interesting suggestion. We are however not really sure on how normal it is to extend this definition into several dimensions and how intuitive this description would be. We did try to use the Backus-Gilbert Spread definition for altitudes, but in the end our conclusion was that since the FWHM is the most common definition used in microwave litterature, and is easy to extend into 2 dimensions we settled on this. Out of curiosity, if the referee has some litterature where the P&H definition is used in multidimensional retrievals we would appreciate a reference for our enlightenment.

Referee # 1 **p1569 l6/7:** *Is it easy to understand that the FWHMs of retrieval error and AVK can be different? Normally, they are different because K is 'non-diagonal' (I know that this is sloppy wording because K is not usualy quadratic; I mean there is cross-talk over altitudes). This makes the retrieval error correlated in the altitude domain even if there is no constraint, i.e. even if the AVK is diagonal. But here the situation is different: In K there is no cross-talk between times, i.e. the only mechanism which can lead to retrieval error correlations in time is the smoothing constraint. Both describe how a variation of the measurement at one time affects the result at another time. It is not obvious to me where the differences come from. Some discussion of this issue would be appreciated. Is it because this is coupled to mutual error propagation in the altitude domain as discussed above?*

The off diagonal elements come, as Referee #1, writes from the a priori constrains which results in several measurements being used to retrieve a single measurement time. We did not want to go into details about all the correlations in the different error matrices, since these are not usually discussed, and as such would need a larger introduction before going into details. This, combined with the fact this discussion is not critical for conveying the message of the paper, means that we find a thorough discussion superfluous. However, to clarify we have added the text: "This correlation is introduced through the temporal correlation in the apriori constrains."

Referee # 1
p1569 bottom: *is there any systematic (even if indirect) dependence of noise with time? If so, noise-weighted averaging might lead to a sampling bias within the 3 hrs intervals. This could be helped by weighting the measurement times accordingly and by assigning the measurement to the weighted mean of the time.*

p1570 l 14/15: *Same issue as above: If noise varies between the 3 hrs intervals, how do you know that it does not vary within the 3 hrs intervals? And if it varies within the 3 hrs intervals, the weighting will cause a temporal sampling bias.*

We agree with Referee #1 that sampling and averaging with a weighted mean can result in a sampling bias, and that this could be alleviated by weighting the measurement times accordingly. However, the measurement data from the OSO radiometer that we work with in this paper is processed this way. To clearly state our assumptions we added the text "This is done since we assume that the thermal noise in the measurements is uncorrelated to the concentration of water vapour during the 3 h interval. " to section 4.1.

The weighted averaging of the 3 h spectra into 48h spectra is mainly done to compare this averaging with the time series inversions. So even if we agree with Referee #1 that applying a similar weight to the measurement times is more accurate we feel we think that this averaging method is the most appropriate for this paper, as the time series inversion will end up doing a similar averaging, and not move the retrieval grids .

Referee # 1 p1570 l3/4: *the term 'retrieval matrices' sounds like a technical term but it is not, which is confusing. I guess you need a generic term including covariance matrices and Jacobians etc. Further, the term 'large' is not very precise. I suggest 'the dimensions of the matrices involved in Eqs. 3-4 become quite large.'*

We agree to this comment and have made the appropriate changes.

Referee # 1 p1570 middle*: I appreciate that the most important parameter settings for the real data retrieval are summarized here but I consider it helpful to additionally refer to a publication where the retrieval details are explained in more depth if available.*

An additional reference to Forkman et al. (2003) and (Haefele et al., 2009) have been added.

Referee # 1 p1571 l6*:... but N is smaller than in the theoretical case (for the same overall time period), since there were no data gaps. So the N+N' corresponds to the N of the synthetic case, doesn't it. I find this somewhat confusing because one might think that the new N+N' is necessarily larger than the old N. Some clarification of this will be appreciated.*

We understand that the wording of these sentences might be a bit confusing, but as defined in section 2.2. N is equal to the number of measurement times, so if we add point to fill data gaps the new size of K and x will increase as described in the text. To clarify this a definition of N is added to the scentence: "..., and *N* the number of times with measurements,...".

Referee # 2: 1573- line 8: *Given the choices made for the errors in the NatMean method, the temporal resolution and the retrieval noise (as shown in Figure 5) actually are quite different from that of the averaging method even at 76km. Hence the results in Figure 6 be- tween the two methods actually look quite different at this altitude; in contradiction to the statement made here. It's unfortunate that the authors did not choose an altitude where the retrieval noise was similar, but they should point the difference and explain it.*

The differences is retrieval noise is exists as the time series method makes an "optimal" compromise between noise, resolution and sensitivity. This will naturally result in a different amount of noise, than for the averaged retrievals. In spite of this, we do believe that the results from the time series method resemble the averaged spectra results at high altitudes, but we agree that using the word similar might lead to confusion. We have therefore changed the sentence to read: "However, the main point of these retrievals is not to determine the true water vapour concentration in the atmosphere, but rather to show that the time series inversions produce comparable results to the single spectrum inversions at lower altitudes while producing results more similar to those of the averaged inversions higher up."

Referee # 2 1573- line 21: *How is the decreased measurement response due to the fitting of base-line polynomials taken into account in the method presented here? Does this mean that off-diagonal Se terms are now included?*

The fitting of the baselines are not done by modifying the Se matrix, but rather by introducing a new state variable in the x vector to the forward model for each measurement. This is the standard way of doing it for microwave instruments, and introducing this fit does not imply that Se has changed, i.e it remains a diagonal matrix. To clarify this the description of how the baselinefit is applied is now expanded in the last paragraph of section 4.1. Additionally we have added to the text that these fits are performed without correlation in time, i.e. no off diagonal elements is the Sa matrix describing these polynomial fits.

Referee # 1 p1576 middle: *again please add a caveat about periodic a priori correlations (diurnal variation etc).*

A caveat about variables and correlations that are hard to describe using a normal distrubtion, e.g. diurnal variation, is added to section 6.

Referee # 1 p1576 l20 and 2*7 Wouldn't it be better to replace 'practical' by 'based on real data'?*

We agree and the manuscript has been changed.

Referee # 1 p1576 l21 'normal desktop computer': *journal articles are written for eternity, not for the moment. What do you think a reader in 10 years from now understands is a normal desktop computer? Please be more specific. Further, non-microwave remote sensing scientists do not know how many spectral gridpoints you use (I suspect you do not calculate radiative transfer on the final output channels but on a finer monochromatic grid), on which altitude grid the retrievals are represented etc, if you do explicite line-by-line calculations or if you use pre-tabulated absorption cross-sections etc. All these choices affect the calculation time. Either mention all these key data, or consider to remove the statement on calculation time.*

The statement about calculation time is removed and replaced with a more general statement about the calulation time compared to the single spectrum inversions.

Referee # 1 Discussion: *please add a caveat that there can be artefacts if the time averaging kernels are not considered by the data user: Imagine something like the tropical tape recorder. Naively analyzing the tropical H2 O tape recorder in your inverted time series, the decrease of its amplitude with altitude would be overestimated etc. I appreciate the benefits of this method, but as every method involving priori information, there are also traps for the data user, which should be honestly discussed.*

This caveat applies to all multidimensional retrievals, but we agree that this complication exists, as such we have added the following sentence to the end of section 6: "However, as the method increases the dimensions of the retrieval and resulting averaging kernels, it also increases the complexity for any end-user using the data. "

Referee #2: Typos:

1557 – line 6: an "are" is missing after "altitudes"
1566- line 24: drop the "s" on "corresponds"
1575- line 12: drop the "s" on "makes"
C537
1577- line 14: drop the "s" on "applications"

We have fixed these typos.

Referee # 2 Figure 6: *What exactly does "H2O [relative]" mean? Based on the fact that the 1-D measurements are more nearly ∼1, I would guess that a value of "1" implies that the retrieval is giving the a priori. Please provide the defining mathematical expression for this.*

In the main text where figure 6 is discussed we have now added the line: "The results of the retrievals at three different altitudes are shown in Fig. 6 in units relative to the apriori (i.e. 1 = apriori concentration). "