

Response to Referee #1: Matthias Schneider; Dec 2018

Black: Referee's comments

Green: Author's reply

We would like to thank Matthias Schneider for his helpful comments and suggestions. We have taken all the comments into account. In our opinion, the revised has improved thanks to suggestions provided by both reviewers.

the manuscript addresses the following topics:

(1) it describes the retrieval of water vapour profile from ground-based FTIR measurements, (2) it compares the retrieval data with frost point hygrometer sonde data, (3) it assess the impact of different WV a priori data on the retrieval of water vapour, and (4) it assess the impact of different WV a priori data on the retrieval of other trace gases where water vapour is an important interfering species.

My general comments:

I find (4) is a nice and valuable demonstration of the importance for using actual WV profile data in order to avoid large uncertainties in the retrievals of other trace gases. The reason is that WV is very variable and not well capturing the variability results in large retrieval errors of the other species. However, this part of the paper could be further improved by inserting references on previous work where the interference error of WV has been calculated.

We appreciate this comment. It was valuable that the reviewer included references (in the specific comments) from previous work. We have included the references in the revised manuscript. Please see also our response in the specific comment provided below.

I think (1)-(3) need revisions. A constrained remote sensing data product (here x_r) means that a priori data (here x_a) has been updated with a measurement. The product (x_r) strongly depends on the a priori data (x_a). In particular if x_a is variable on small scales (like for WV) the variability in x_r will, to a large extent, reflect the variability of the prescribed x_a . Instead of assessing the quality of x_r the authors should assess how the remote sensing measurement can improve the assumed a priori state of the atmosphere, i.e. the authors should validate $x_r - x_a$ by comparing it to $A(x_s - x_a)$, where x_s is the FPH reference.

This point is well taken and along with suggestions from reviewer #2. In the revised manuscript we have included comparison of retrieved WV and sonde FPH using the formalisms by Rodgers and Connor (2003). Figures 5 and 6 include smoothed FPH profiles using equation 4 in Rodgers and Connor (2003), instead of showing them in the supplemental material (like the initial version). Additionally, we have included in the main text, results of comparisons between FTIR retrievals for both un-smoothed & smoothed FPH. We kindly refer the reviewer to the revised manuscript for additional/modified text and figures. In particular, table 3 summarizes the findings of both comparisons. Furthermore, in the revised section 4.3 we added a short description regarding the value of $x_r - x_a$. Please see our response in the specific comments below.

Furthermore, when using an a priori data that already captures most of the variability, the solution state should be more constrained (the S_a matrix should have much smaller entries) than when using an x_a that captures only few variability. However, judging from Sect. 3 it seems that the authors use a single S_a for constraining the different retrievals.

This has been addressed in the specific comment below.

Specific comments:

I have inserted my ideas/suggestions in the attached pdf version of the manuscript.

Best regards.

Comments provided by the referee are copied from the pdf and shown below in back.

Summary of Comments on amt-2018-283_MS.pdf

Page: 2

Number: 1 Author: pa5682 Subject: Cross-Out Date: 12/3/2018 8:01:36 PM

Accepted, text has been removed.

Number: 2 Author: pa5682 Subject: Inserted Text Date: 12/3/2018 7:55:01 PM
uses

Accepted.

Number: 3 Author: pa5682 Subject: Inserted Text Date: 12/3/2018 10:25:15 PM
/FTIR spectra measured at 12 different sites

Included with minor edits.

Number: 4 Author: pa5682 Subject: Inserted Text Date: 12/3/2018 10:36:31 PM
for generating a long-term data set of global representativeness of tropospheric water vapour profiles with a DOFS of almost 2.8 and of about 1.6 for the ratio between the most abundant isotopologue H216O and the heavy isotopologue HD16O.

Included with minor edits.

Number: 5 Author: pa5682 Subject: Inserted Text Date: 12/3/2018 8:03:40 PM
Comparisons of FTIR and operational radiosondes have been used to validate optimized WV profile retrieval strategies (e.g. Schneider et al. 2006; Schneider and Hase, 2009; Schneider et al., 2016).

Included with minor edits.

The complete paragraph now is:

MUSICA (MULTi-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water) is a project within the NDACC/FTIR using standard spectra from a subset of NDACC sites in order to generate a long-term data set of tropospheric water vapor profiles with degrees of freedom (DOF) of about 2.8 and of about 1.6 for the ratio between the most abundant isotopologue $H_2^{16}O$ and the heavy isotopologue (Schneider:2012, 2016, Barthlott, et al., 2017). Comparisons of FTIR and operational radiosondes have been used to validate optimized WV profile retrieval strategies, (Schneider et al., 2006; Schneider and Hase, 2009). Vogelmann et al. (2015) studied the spatial-temporal variability of WV in the free troposphere (Zugspitze, Germany) by exploiting the geometry of measurements of differential absorption lidar (DIAL) and FTIR. In particular, they assessed the variability in short scales, i.e., few kilometers and minutes.”

Number: 6 Author: pa5682 Subject: Inserted Text Date: 12/3/2018 10:37:43 PM
made at two different sites, that have so far not been considered within MUSICA. We use spectral microwindows that are not identical to those of MUSICA (Barthlott et al., 2017, Fig. 1 therein) and perform the inversion on a linear scale (instead of a logarithmic scale used by MUSICA

Thanks for pointing this out. Rather than in the introduction these details have been included in Sect 3 (Retrieval of water vapor from FTIR). The following sentence is now in Sect 3:

“We use spectral micro-windows that are not identical to those of current MUSICA version (Barthlott et al., 2017) and perform the inversion on a linear scale (instead of a logarithmic scale used by MUSICA).”

Page: 5

Number: 1 Author: pa5682 Subject: Highlight Date: 12/7/2018 7:19:46 PM
For constraining the authors use the same S_a^{-1} for the different a priori from Section 4.3?
If yes, the solution state will be much looser constraint for the daily varying a priori than for the monthly varying a priori, i.e the two retrievals are difficult to be compared.

In principle, we agree that S_a might need an adjustment depending on the a priori, especially for gases with less variability. The variability of water vapor can be large and not necessarily the S_a of the daily a priori is always much looser than the monthly profile. There are cases where even the daily a priori is distant from the “real” or retrieved WV but the monthly a priori is better. Optimizing a S_a for one day might not necessarily be best for another day. Furthermore, there are several reasons why we use a single S_a : (1) this work does not aim to optimize retrieval parameters but rather use a common retrieval approach that could be applicable to more sites; (2) the standard S_a used here has been optimized in all cases in order to obtain similar information content; (3) three out of four a priori are similar and we expect similar S_a and changing S_a would result in a more complex comparison.

We value this comment and in the revised manuscript we include this modified text (please note that it was modified also following the suggestion of reviewer #2):

“The S_a matrix is specified at each layer as a fraction of the a priori profile, which allows for a linear scaled retrieval. We adopted a maximum variability of 50 % in the diagonal covariance and exponentially decreasing by altitude. In order to prevent sporadic vertical profile oscillations, we include a Gaussian correlation length of 25 km in the off-diagonal elements of S_a . This S_a has been optimized in order to obtain similar information content for all a priori presented in section 4.3, a requirement for efficient processing of decades of NDACC spectra.”

In addition, in the conclusions we mentioned that further optimization of the a priori covariance matrix might be needed in future research. The paragraph reads as:

“Further research would explore the additional WV absorption features in order to improve the information content, e.g., micro-windows employed in the latest MUSICA version. Also, as we show, the ERA-I WV profiles yield lower biases, hence we would construct a priori covariance matrices for these that maximize accuracy and vertical structure.”

Page: 6

Number: 1 Author: pa5682 Subject: Highlight Date: 12/7/2018 6:24:45 PM

It is important to describe here the remote sensing measurement as an update of the a priori information, i.e. the actual measurement is $x_r - x_a$ not x_r ! The authors should explain how the a priori affect the retrieval results: $x_r = A(x - x_a) + x_a + D_x$, where D_x are retrieval errors. Because A is not an identity matrix x_r will always significantly depend on x_a .

We include a description of the effect of the apriori in Section 4.3:

“The optimal estimation method is influenced by the a priori profile because it may bias the solution of equation 1. Since WV is highly variable, even in time scale of hours, using the most accurate a priori might improve the retrieval results. In general, the retrieval of WV can be seen as an update of the a priori information.”

The authors use variable a apriori data (see Sect. 4.3) and the variability seen in the retrieved vertical profile reflects to large extent the variability prescribed by the a priori data.

Unfortunately this is not correctly considered in Sect 4 of the paper. Because the authors work with a variable a priori they should evaluate the signals in $x_r - x_a = A(x - x_a)$, because this is the measured signal not $x_r = A(x - x_a) + x_a$!. Furthermore, the authors assume that if x_s can be used as a reference for the retrieved profile x_r ; however, actually x_s is highly resolved and absolutely calibrated reference, i.e. it is a reference for the real atmospheric profile x . So correctly the authors should compare $x_r - x_a$ with $A(x_s - x_a)$ in order to validate the remote sensing measurement.

In the revised section 4.3 we added a short description regarding the value of $x_r - x_a$ to evaluate signal of the measurements:

“Additionally, the difference between WV retrievals and a priori profiles (x_r-x_a) provides further evidence in the measured signal and to some extent the variability prescribed by the a priori (Rodgers and Connor, 2003). For example, this difference is about 11 +/- 38 % using ERA-6 while for WACCM is about 29 +/- 32 % for the first layer. As we expected, from these observations it can be seen that the WACCM climatology as a priori results in greater deviations compared to ERA-6.”

As pointed out before for the comparison we follow the formalism of Rodgers and Connor (2003) in addition to the un-smoothed comparisons to assess vertical gradients and avoid averaging kernels limitations.

Page: 12

Number: 1 Author: pa5682 Subject: Highlight Date: 12/4/2018 2:10:17 PM

The retrievad data x_r are almost not sensitive to atmospheric variations above 10km, so I suggest not showing layers above 10km.

We have removed the last two layers and text has been modified accordingly.

Page: 13

Number: 1 Author: pa5682 Subject: Highlight Date: 12/4/2018 2:14:12 PM
monthly varying a apriori is used, please specify?

“a 40 year simulation (1980-2020) of the WACCM mean profiles”

Number: 2 Author: pa5682 Subject: Highlight Date: 12/4/2018 2:14:56 PM
daily varying a priroi?

“daily varying (ERA-d)”

Number: 3 Author: pa5682 Subject: Highlight Date: 12/4/2018 2:15:31 PM
6 hourly varying a priori profile?

“6 hourly varying WV vertical profiles (00, 06, 12, and 18 UTC) obtained from ERA-I (ERA-6)”

Number: 4 Author: pa5682 Subject: Highlight Date: 12/4/2018 2:16:28 PM
daily varying a priori profile

“daily varying NCEP/NCAR (NCEP-d) reanalysis WV profiles”

Number: 5 Author: pa5682 Subject: Highlight Date: 12/4/2018 2:28:19 PM
If x_s is the reference $x_r-x_s=A*(x_s-x_a)-(x_s-x_a)+Dx=(A-I)*(x_s-x_a)+Dx$
 Dx are the retrieval errors

This is correct but we do not have anything to add.

Page: 14

Number: 1 Author: pa5682 Subject: Highlight Date: 12/4/2018 2:36:40 PM

$x_r - x_s = (I - A) * (x_a - x_s) + Dx$, i.e. it depends on the retrieval errors Dx and on $(x_a - x_s)$.

Because a daily or even 6 hourly varying x_a better captures the actual variability of atmospheric WV, using x_a from ERA-d and ERA-6 (instead of monthly climatologies) better captures the variability as given in x_s , i.e. $x_r - x_s$ shows a particular small scatter. This is no surprise.

In general, this might be true, but in this work, we present a quantitative assessment of the different a priori, even for daily or 6 hourly.

Page: 15

Number: 1 Author: pa5682 Subject: Highlight Date: 12/3/2018 10:59:53 PM

see comment on Table 2

ok

Number: 2 Author: pa5682 Subject: Highlight Date: 12/3/2018 11:39:37 PM

Here retrieval data generated by using varying a priori data are compared to the sonde measurements. This does not allow robust conclusions on the quality of the FTIR measurements. Actually there will be a very good agreement already by comparing the varying a priori with the sonde measurement (i.e. without any information from the FTIR measurement).

What you need to compare and evaluate is the difference with respect to the a priori! So you have to calculate $x_r - x_a$ and correlate it to $A * (x_s - x_a)$.

In the revised table we have added the comparison using the formalisms by Rodgers and Connor (2003), i.e., smoothing the FPH profiles using the water vapor averaging kernels. Note that results using “un-smoothed” are also shown. The limitations of the averaging kernels are seen clearly in this table.

Page: 16

Number: 1 Author: pa5682 Subject: Highlight Date: 12/5/2018 3:03:57 PM

There is some work:

The impact of interferences from WV on the retrieval of CO has been estimated by Sussmann and Borsdorff, 2007 (doi:10.5194/acp-7-3537-2007), on the retrieval of O3 by García et al. 2014 (doi:10.5194/amt-7-3071-2014) and on the retrieval of CH4 by Sepúlveda et al., 2014 (doi:10.5194/amt-7-2337-2014).

Thanks for providing the references of previous works. The text in the revised manuscript has been edited to include the references provided. We note that García et al. (2014) and Sepúlveda et al. (2014) retrieved water vapor in a first step to minimize errors in the retrieval of O₃, and CH₄, respectively. Sussmann and Borsdorff. (2007) quantified the impact of water vapor in the retrieval of CO and further apply a retrieval strategy to remove interference errors. Still, the effect of using co-located and highly-resolved WV are missing in the literature.

Page: 17

Number: 1 Author: pa5682 Subject: Highlight Date: 12/7/2018 6:45:22 PM

The strategy is sufficient to avoid WV interferences in the retrievals of other trace gases and the obtained WV profiles are of a reasonable quality. However, using retrievals with more WV lines, and retrievals on log scale (tropospheric water vapour is log-normally distributed) should theoretically provide better results. Actually the MUSICA WV data show higher DOF and agreement with radiosonde also within 10-20%

The above suggestion has been included and reads as follow:

“This example suggests that the current retrieval strategy of WV is suitable to avoid WV interference in the retrievals of other trace gases.”

In addition, in the conclusions we included the following:

“Further research would explore the additional WV absorption features in order to improve the information content, e.g., micro-windows employed in the latest MUSICA version. Also, as we show, the ERA-I WV profiles yield lower biases, hence we would construct a priori covariance matrices for these that maximize accuracy and vertical structure.”

Response to Anonymous Referee #2; Dec 2018

Black: Referee's comments

Green: Author's reply

We would like to thank Reviewer #2 for taking the time to review our manuscript. We believe the revised manuscript has improved thanks to thorough and thoughtful comments provided.

1 General Comments

The paper deals with two main topics:

- a) first the authors assess the limitations in retrieving the real Water Vapor (WV) vertical variability from the boundary layer to the upper troposphere - lower stratosphere, with a standard inversion of FT solar absorption measurements in the middle-infrared. The study includes the validation of WV profiles retrieved from round based FTS measurements operated from Boulder (Colorado) and Mauna- Loa (Hawaii), via intercomparison with WV profiles measured by state-of-the-art Frost Point Hygrometers (FPH) operated from balloons.
- b) Secondly, a sensitivity study is presented, showing the error on retrieved HCN, CO, and C2H6 VMRs due to assuming a less than perfect WV vertical profile.

The subject of the paper is clearly within the scope of AMT. The methods used are scientifically sound, the presentation is sufficiently concise, however it could be improved by rephrasing a few sentences as outlined in the specific comments reported below. The paper does not introduce novel concepts or ideas, however the results of the study will be useful for other scientists using the data presented or data deriving from similar measurements. For this reason I recommend this paper for publication in AMT, after some revisions as outlined below.

My main comment or criticism is about the strategy the authors adopt to deal with the Averaging Kernels (AKs). I agree that the AKs may not be a sufficiently accurate tool to evaluate the smoothing error of the retrieved WV profiles. This is due both to the fact that AKs are only a "linear" approximation of the vertical response function of the measuring system (instrument plus retrieval algorithm), and to the fact that it is generally hard to setup a covariance matrix which represents properly the variability of WV from ground to the Upper-Troposphere / Lower Stratosphere (UTLS). To show the limitations (in your test case) of the smoothing error as derived from AKs and the Rodgers (2000) approach, rather than moving the AKs analysis to the supplemental material, I would have compared, in the main paper, the actual smoothing error (obtained via intercomparisons with FPH) with the estimate of the same error obtained from AKs.

We agree that adding the comparison of smoothed FPH profiles in the main text improves the quality of the manuscript. In the revised manuscript we have included smoothed FPH profiles in Figures 5 and 6, instead of showing them in the supplemental material. Additionally, we have included in the main text results of comparisons between FTIR retrievals with both un-smoothed and smoothed FPH. We kindly refer the reviewer to the revised manuscript for

additional/modified text and figures. In particular, table 3 summarizes the finding of both comparisons.

The second general comment I have is about the sensitivity analysis presented in Sect. 5. To my opinion it would be worth to better explain why, after the analysis presented in the first part of the paper, then you start to study a quite different subject, such as the mapping of WV errors on subsequent VMR retrieval of other gases.

While the first section of the paper is focused in the retrieval/comparison of water vapor, we do not consider that section 5 is completely different. The first section considers FPH water vapor as reference and the water vapor sensitivity in the retrieval of other gases also considers FPH as reference, considering that is rare to have fully resolved coincident measurements we believe this is the right place to show both. We do agree that further details might be needed to better explain this second section. In the revised manuscript we slightly have expanded the description of this second part, mainly in the introduction.

More-over, since your measurements cover the middle-infrared, I also expect a sensitivity of the retrieved VMRs to the temperature error. This is already shown in Fig. 3 for WV. What about the error on HCN, CO, and C₂H₆ VMRs due to the temperature error ? Do you suggest to retrieve also the temperature profile from the same measurements or you are satisfied with the temperature profiles taken from NCEP at NDACC ?

As shown in the manuscript the importance of pre-retrieving water and use it in the retrieval of other gases is important but can be gas and site specific. The effect of using a daily NCEP temperature profile versus a more refined temporal temperature profile (or joint gas/temperature profile) can also be site/gas specific. Similar as WV, full error analysis is considered HCN, CO, and C₂H₆, i.e., temperature profile uncertainty is considered. In the revised manuscript we added a sentence explaining this.

Previous studies, e.g., Schneider and Hase (2008); Schneider et al. (2008), have shown that a joint retrieval of temperature profiles significantly improves the quality of O₃. In general, we do expect similar results for other gases. The temperature sensitivity is out of the scope of the manuscript but have added a brief description of the joint approach from previous studies in the revised manuscript.

In addition, just for reference, below there is a description of our current approach in the estimation of the error in the temperature profiles.

Errors in the temperature profile can have both systematic and random components. In our sites, we quantify these components by comparing radiosonde temperature profiles at or near each site with the daily NCEP temperature profiles. Both radiosonde and NCEP temperature profiles are interpolated onto the retrieval input grid for each site. The mean and the standard deviation of the differences between the NCEP and radiosonde temperature profiles are calculated. The mean of the difference can be as the systematic component of the error, while the standard deviation of the difference can be viewed as the random component. This is carried out for several years. Then, we include this onto the forward model parameters errors to obtain the covariance matrix

for the forward model parameters. Ideally, the same can be done for WV, however high quality and coincident measurements of WV for all sites is rare, hence the importance of the sensitivity analysis given in the second part of the manuscript.

2 Specific Comments

P4 L27,28: Constraining is important to select the solution which, among the possible solutions of the ill-posed inversion, is the most likely on the basis of our prior knowledge.

We edited text and included the reviewer's suggestion.

P4 Eq.1: Here it is not clear if your retrieval performs only a single or several Gauss-Newton iterations, because you don't have an iteration index in the Equation. Please, also define clearly the meaning of K. Do you compute it at each iteration? I guess K is the Jacobian of the forward model with respect to the retrieval parameters, therefore it should be re-computed at each iteration and should show an iteration index.

Thanks for catching this up. We have modified the equation and text accordingly.

P5 L2: It would be interesting here to know which is the spectral range covered by the spectrometers used, and which is the rationale behind the selection of the listed micro-windows (e.g. minimum retrieval error?).

Measurements at BLD and MLO follow standard measurement protocols of the InfraRed Working Group (IRWG) of NDACC. Several optical band pass filters are used to maximize the signal to noise in the middle and near infrared ($\sim 700 - 5000 \text{ cm}^{-1}$), described briefly in section 2.1. As mentioned in the manuscript, we do not aim to optimize a retrieval strategy of WV but rather to use a retrieval approach that we have been following in the past year. Past studies have shown that multiple micro-windows can be used. For example, Schneider et al. (2006) applied signatures between $700-1400 \text{ cm}^{-1}$; Schneider et al. (2012) used several micro-windows to retrieve water vapor isotopologues for the FTIR MUSICA retrievals. In particular for WV they used the $\sim 2800 - 2900 \text{ cm}^{-1}$ range. Later, Barthlott et al. (2017) removed strong micro-windows used in Schneider et al. (2012) with strong absorption and added weaker lines in the 2700 cm^{-1} range. An extensive number of micro-windows can be used to retrieve water. In our case, the micro-windows we use cover a similar range ($\sim 2800 - 2840 \text{ cm}^{-1}$) and were tested to maximize the information content and minimize total error and they give consistent results across a wide range of WV columns. We added the following sentence in section 3: "*These micro-windows have been chosen to maximize the information content and minimize total error.*"

P5 L8,9: I got an idea of what the authors would like to say, however I suggest to re-phrase more clearly this sentence.

After reviewing section 3, we believe the following sentence:

“The SNR determines how much influence the spectra has in each micro-window versus the a priori, as well as to characterize the measurement error described in section 3.1” can be removed since the SNR is primarily used in the error analysis section 3.1.

P5 L10: Which are the “relaxed covariance matrices” that induce oscillations ? Please clarify.

The following paragraph:

“In order to prevent sporadic vertical profile oscillations due to relaxed covariance matrices we implement ad hoc diagonal elements of Sa with a maximum variability of 50% at the surface and exponentially decreasing by altitude with a inter-layer thickness correlation coefficient. A Gaussian correlation with a length of 25 km is used for the off-diagonal elements of Sa . ”

has been replaced by:

“The Sa matrix is specified at each layer as a fraction of the a priori profile, which allows for a linear scaled retrieval. We adopted a maximum variability of 50 % in the diagonal covariance and exponentially decreasing by altitude. In order to prevent sporadic vertical profile oscillations, we include a Gaussian correlation length of 25 km in the off-diagonal elements of Sa . This Sa has been optimized in order to obtain similar information content for all a priori presented in section 4.3, a requirement for efficient processing of decades of NDACC spectra. ”

P5 L14: If apodization is not used, how broad is the ILS used in the forward model to emulate the instrument effect ?

The spectra are recorded at an OPD of 250 cm giving an unapodized spectral resolution of nominally 0.004 cm^{-1} . At this resolution all atmospheric spectral features are fully resolved in this MIR region. We have specified this OPD in the text to be clear.

P5 L25,26: Here it is not clear how the 0.5% rms error “on the fit” maps onto the retrieved WV. Does this rms error refer to the “residuals of the fit” or directly to the retrieved WV ?

To make it clear, the description of this has been modified accordingly. The revised paragraph is:

“We examined the effect of using more temporally refined temperature profiles. In general, the six hourly temperature profile from the ERA-I reanalysis model, produced by the European Center for Medium-Range Weather Forecasts (ECMWF) (Dee, et al., 2011), follows the daily average temperature profile shape very well for both sites. The root mean square error (rmse) between the six hourly data of ERA-I and daily average temperature is less than 0.5% using 2013 data for both BLD and MLO and the biases are less than 0.25 % for BLD and less than 0.1 % for MLO. These results suggest daily mean temperature should be adequate for retrievals but we further investigated the sensitivity of water vapor to this variability and found that water vapor agrees within 1 % if using the daily average profile. The temperature profile uncertainty is considered in the error analysis in section 3.1.”

P6 Fig.1: It would be better to show CH4 and N2O absorption contributions with different line colours.

In the updated figure species are identified with the same color for the different micro-windows.

P6 L14-16: This bias is with respect to the a-priori state vector which, in turn, will probably have some bias with respect to the real profile. Note that if the bias of the retrieval was known both in sign and amplitude, then it would be possible to correct for it...

The text has been updated with the above description.

P7 L1-5: Here I would state clearly which is your retrieval vector. Do you retrieve a WV profile using the discretization mentioned in the colour scale of Fig.2a ? Do you include further fitting parameters ? (Such as atmospheric continuum, for example).

The color code scale in Fig. 2a illustrates the kernels on the retrieval grid in the troposphere but the retrieval grid does go up to 120 km. Atmospheric continuum is not considered.

P7 L7-9: Smoothing the high-vertical-resolution profiles via the averaging kernels of the coarse-vertical-resolution experiment is not mandatory, especially if you attribute a “smoothing error” to the profile differences or if you want to try characterizing the smoothing error itself. Therefore I would simply state your choice here, without trying to find a justification, which is also rather fuzzy to my view.

This description has been removed and we further improve this description in section 4. We kindly refer the reviewer to the updated text, especially in section 4.

P7 L18: Off-diagonal elements of Se may play a very important role if the spectrum is oversampled (wrt interferogram) and/or if apodization is used. Please state explicitly that, apparently, this is not your case.

In the case of very high spectral resolution spectra that fully resolve all absorption features where no apodization is used off diagonal elements are in general rendered moot. The interferogram processing may zero fill for speed but no excessive zero filling is used.

P8 L8: Please define also the symbol Kb .

Done

P8 L17-ff: In Fig.3a the error due to interfering species is also shown. Which are the considered interfering species that are not simultaneously retrieved with WV ?

The interfering species are those simultaneously retrieved. We wish to take into account the uncertainty in those lines on the retrieval of the target species. This is typically a small contributor.

P9 L1: This sentence is not clear and may be questionable. Does this mean that your AKs are not a good estimate of the vertical response function of your system (instrument plus inversion scheme) ? Why ?

The paragraph of this section has been updated, based on previous suggestions. As mentioned before the revised manuscript shows smoothed FPH profiles and quantitative comparisons of FTIR with both un-smoothed and smoothed FPH profiles. In general, the biases are lower when compared with un-smoothed profiles pointing out that FTIR AKs may not be adequate tool to use in the comparisons. This is pointed out in the updated conclusions.

P9 L7: Please note that re-gridding via interpolation is, on its own, an arbitrary smoothing. So, I do not fully understand why you do not want to use AKs to smooth and re-sample high-resolution profiles prior to intercomparison (as it seems you already did in the plots presented in the supplemental material).

The revised text of this section includes the comparison with smoothed profiles, as suggested in the first general comment. We kindly refer the reviewer to the updated text

P9 L12-14: As shown in Fig. 2c, the FTS retrievals have less than 3 DOFs therefore, why using so many layers for the intercomparison ? The risk is to find biases of different sign in adjacent layers.

This might be true for most gases. In fact, we do normally use total and/or partial columns depending on the number of independent pieces of information. However, a goal of this paper is to assess the ability to retrieve water vapor gradients. Hence, we decided to use these layers. Furthermore, as stated before the averaging kernels might not be a proper tool to assess sensitivity.

P10 Sect. 4.1: The underlying idea is good, however, please note that the variability evaluated here could underestimate the real WV variability, due to the constraint of the retrievals towards the a-priori state vector. I would have estimated from measurements the variability of the spectrum vs time and would have derived the corresponding “time- mismatch” error covariance matrices relating to the individual WV profiles, using Eq. 3. I suggest to include a comment on this regard (or change approach...).

In the manuscript is already stated that the real variability might be greater because of potential lost variability during retrieval smoothing. In general, we see the stability of the spectrum by means of the signal to noise ratio (SNR), which does not change significantly, however we see changes in WV spectral features so we see a change in amount of WV. Note that we also aim to see variability of water vapor at several altitude layers and analyzing the spectrum would not give us that information. In the revised manuscript we have removed the last two layers since we do not have enough sensitivity.

P11 Sect. 4.2: Here I did not understand if, from this analysis, you also derive an estimate of the error component to be attributed to the difference between FTS and sonde WV profiles, due to

the spatial mismatch of the measurements. I agree that it is hard to derive such an error estimate however, lacking this estimate, I do not see very much the usefulness of this section. Please explain.

As described in this section, we aim to assess the spatial mismatch between the sonde at various altitudes and the maximum sensitivity location of the FTIR. This assessment is already complex, and to our knowledge the first time applied at various altitudes. We actually believe contains great value because as mentioned at the end of this section the spatial mismatch depends on the complex convective dynamics and not only in the coincidence time interval. It is already mentioned that a thorough assessment of the spatial variability would require measurements of an extensive area simultaneously and at different altitudes. However, in the revised text we are explaining further that an error due to spatial mismatch is not derived and only an assessment of the spatial mismatch is aimed.

P13 L2,3: The second effect of a-priori WV on the solution is not clear. Did-you mean that the a-priori WV influences the solution also because it is used as initial guess for the Gauss-Newton iterations ? (This latter effect should be negligible if the retrieval converges properly).

We have re-phrased the sentence as follow:

“The optimal estimation method is influenced by the a priori profile because it may bias the solution of equation 1.”

P16, Sect. 5: the link between this Section and the work presented in the previous Sections of the paper is non very clear. Maybe you could state at the beginning of this Section how the work you are going to present is linked with the analysis presented earlier.

The first paragraph of this section has been modified in order to make clear the connection with previous findings.

P16 L1,2: What is the meaning of “expecting WV” in this context ? Usually an optimized microwindow selection scheme tries to avoid spectral interferences from WV and re- lated isotopologues. From the second sentence, however, I understand the opposite.

“expected” has been replaced by “present” in the revised text.

P17 Table 3: Which is the rationale for the adopted sorting of interfering species in the Table ? Perhaps their relative importance ? In this case one should assess the retrieval errors due to the interference of ozone. In this spectral region I also expect temperature knowledge to be of importance (see general comments above). Moreover, I understand that the micro-windows used were selected in already publishes papers, however you could at least mention here the rationale with which they were selected / optimized (e.g. with the aim of minimizing the total retrieval error of the gas to be retrieved).

We added the following sentence in the revised text:

“Table 3 presents the interfering species with strong and/or weak absorption signatures within each micro-window for all target gases. In all cases, the selected settings have been chosen in order to maximize the information content and minimize the total error in the retrieval.”

Furthermore, the uncertainty due to spectroscopic absorption of ozone and other species (interfering species) is considered for the final error analysis. Since this is described for water vapor, we just added the following sentence:

“Similar as WV, full error analysis is performed, i.e., mainly considering measurement noise error and forward model parameter error (see Sect. 3.1).”

Please see also our comment above regarding the temperature and error analysis.

3 Technical Corrections

P2 L26: Remove double comma between “Zugspitze” and “Germany”.

Double comma removed.

P4 L3: Description... described elsewhere (rewriting needed).

The sentence has been changed to:

“A thorough description of the FPH measurement technique has been described in Hurst et al. (2011b) and Hall et al. (2016).”

P7 L3: but as explained before, one of the goals...

Done

P11 Caption of Fig. 5: Same as Fig. 4 but for MLO.

Corrected.

P12 L9: As mentioned above, the initial spatial difference...

Corrected.

P13 L17: I do not see the usefulness of putting some figures in the supplemental material when these are recalled and described in the text of the main paper. I would put all the figures in the main paper file (of course only if this operation does not cost too much!)

As mentioned above, two of the initial supplemental figures have been removed (since the smoothed profiles have been included the main text). We decided to keep three figures in the supplement for at least two reasons: (1) we would like to provide a complete and clear process

during the analysis; (2) these images are not critical part in the main text, but of course they are mentioned in the main text so the reader can check them out if they want.

P16 Fig. 9: in the vertical axis labels “ppmv” has a small “v”, is this an intentional choice ?

The subscript has been removed.

References

Barthlott, S., Schneider, M., Hase, F., Blumenstock, T., Kiel, M., Dubravica, D., García, O. E., Sepúlveda, E., Mengistu Tsidu, G., Takele Kenea, S., Grutter, M., Plaza-Medina, E. F., Stremme, W., Strong, K., Weaver, D., Palm, M., Warneke, T., Notholt, J., Mahieu, E., Servais, C., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Tropospheric water vapour isotopologue data (H^{16}O , H^{18}O , and HD^{16}O) as obtained from NDACC/FTIR solar absorption spectra, *Earth System Science Data*, 9, 15–29, doi:10.5194/essd-9-15-2017.

Schneider, M. and Hase, F.: Technical note: Recipe for monitoring of total ozone with a precision of 1 DU applying mid-infrared solar absorption spectra, *Atmos. Chem. Phys.*, 8, 63–71, 2008, <http://www.atmos-chem-phys.net/8/63/2008/>.

Schneider, M., Redondas, A., Hase, F., Guirado, C., Blumenstock, T., and Cuevas, E.: Comparison of ground-based Brewer and FTIR total column O₃ monitoring techniques, *Atmos. Chem. Phys.*, 8, 5535–5550, 2008.

Tropospheric water vapor profiles obtained with FTIR: comparison with balloon-borne frost point hygrometers and influence on trace gas retrievals

Ivan Ortega¹, Rebecca Buchholz¹, Emrys Hall^{2,3}, Dale Hurst^{2,3}, Allen Jordan^{2,3}, and James W. Hannigan¹

¹Atmospheric Chemistry Observations & Modeling, National Center for Atmospheric Research, Boulder, Colorado, USA

²Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA

³NOAA Earth System Research Laboratory, Global Monitoring Division, Boulder, Colorado, USA

Correspondence to: Ivan Ortega (iortega@ucar.edu); James W. Hannigan (jamesw@ucar.edu)

Abstract. Retrievals of vertical profiles of key atmospheric gases provide a critical long-term data record from ground-based Fourier Transform InfraRed (FTIR) solar absorption measurements. However, the characterization of the retrieved vertical profile structure can be difficult to validate, especially for gases with large vertical gradients and spatial-temporal variability such as water vapor. In this work, we evaluate the accuracy of the most common water vapor isotope (H_2^{16}O , hereafter WV)

5 FTIR retrievals in the lower and upper troposphere - lower stratosphere. Coincident high-quality vertically resolved WV profile measurements obtained from 2010 to 2016 with balloon-borne NOAA Frost Point Hygrometers (FPH) are used as reference to evaluate the performance of the retrieved profiles at two sites: Boulder, Colorado and in the mountain top observatory of Mauna Loa, Hawaii. For a meaningful comparison, the spatial-temporal variability has been investigated. We present results of comparisons between FTIR retrievals with un-smoothed and smoothed FPH profiles to assess WV vertical gradients. Additionally, we evaluate the quantitative impact of different a priori profiles in the retrieval of WV vertical profiles using un-smoothed comparisons. An orthogonal linear regression analysis shows the best correlation among all tropospheric layers using ERA-Interim (ERA-I) a priori profiles and biases are lower for un-smoothed comparisons. In Boulder, we found a negative bias of $0.02 \pm 1.9\%$ and precision of 3.7% (r = 0.95) for the 1.5 - 3 km layer. A larger negative bias of $11.1 \pm 3.5\%$ and precision of 7.0% (r = 0.97) was found in the lower free troposphere layer of 3 - 5 km (r = 0.97) attributed to rapid vertical change of WV, which is not always captured by the retrievals. The bias improves in the 5 - 7.5 km layer (1.0 \pm 5.3\%) and the precision worsens to about 10% (r = 0.94). The bias remains at about 13% and the precision remains to about 10% for layers above 7.5 km but below 13.5 km. At MLO the spatial mismatch is significantly larger due to the launch of the sonde being farther from the FTIR location. Nevertheless, we estimate a negative biases of $5.9 \pm 4.6\%$ (r = 0.93) for the 3.5 - 5.5 km layer (r = 0.93) and $9.9 \pm 3.7\%$ (r = 0.93) for the 5.5 - 7.5 km layer (r = 0.93), and positive biases of $6.2 \pm 3.6\%$ (r = 0.95) for the 7.5 - 10 km layer (r = 0.95), and 12.6% and greater values above 10 km. The agreement for the first layer is significantly better at BLD likely that the air masses are similar for both FTIR and FPH. Furthermore, for the first time we study the influence of different sources of WV a priori profiles in the retrieval of selected gas profiles. Using NDACC standard retrievals we present results for hydrogen cyanide (HCN), carbon monoxide (CO), and ethane (C_2H_6) by taking NOAA FPH profiles as the ground-truth and evaluate the impact of other WV profile sources profiles. We show that the effect is minor for C_2H_6 (bias <

0.5 % for all WV sources) among all vertical layers. However, for HCN we found significant biases between 6 % for layers close to the surface to 2 % for upper troposphere depending on WV profile source. The best results (lowest bias/reduced bias and precision and r-values closer to unity) are always found for pre-retrieved WV. Therefore, we recommend to first retrieve WV to use in subsequent retrieval of gases.

5 1 Introduction

Water vapor is an ubiquitous atmospheric constituent with an extremely important role in the lower and middle troposphere and stratosphere: it is the most variable and critical greenhouse gas (Kiehl and Trenberth, 1997); it plays a key role in atmospheric chemistry, e.g., heterogeneous chemistry, aerosol formation, and wet deposition (Seinfeld and Pandis, 2006); it affects global radiation through cloud formation (Dessler, 2011); and acts as the main source for precipitation in the lower atmosphere (Trenberth and Asrar, 2014). Middle and upper tropospheric and lower stratosphere stable water vapor isotopes are key to understanding the water cycle feedbacks such as mixing of air masses, dehydration pathways, and free-tropospheric moisture (Noone, 2012; Galewsky and Rabanus, 2016).

Obtaining consistent long-term observations of vertical distributions of water vapor is challenging but highly desirable in order to understand climate evolution and feedback effects (Held and Soden, 2000). There is a need to measure water vapor vertical distribution for long-term monitoring but there are only few data sets, e.g., in-situ balloon observations in Boulder, Colorado, USA are the longest data set of the most common water vapor isotope ($H_2^{16}O$, hereafter H₂O or WV) with information from lower to middle stratosphere (Oltmans et al., 2000; Hurst et al., 2011b). It has been shown that ground-based Fourier transform infrared (FTIR) measurements provide reliable long-term and continuous observations of WV (Sussmann et al., 2009; Schneider et al., 2010). FTIR measurements have focused mostly on integrated WV analysis among the Network for Detection of Atmospheric Composition Change (NDACC, see <http://ndacc.org>). For integrated WV (IWV, i.e., total columns) FTIR have been shown to be very precise with about 2.2 % using FTIR side-by-side inter-comparisons (Sussmann et al., 2009).

~~The retrieval of WV vertical MUSICA (MUlti-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water) is a project within the NDACC/FTIR using standard spectra from a subset of NDACC sites in order to generate a long-term data set of tropospheric water vapor profiles with degrees of freedom (DOF) larger than two are achieved but there is a lack of comprehensive comparisons of vertical gradients using FTIR with well-characterized highly-resolved independent measurements from the lower to upper troposphere and stratosphere of about 2.8 and of about 1.6 for the ratio between the most abundant isotopologue $H_2^{16}O$ and the heavy isotopologue $HD^{16}O$ (Schneider et al., 2012, 2016; Barthlott et al., 2017).~~
Comparisons of FTIR and operational radiosondes have been used to validate optimized retrieval strategies, e.g., for WV in the upper troposphere/lower stratosphere see Schneider et al. (2006) and Schneider and Hase (2009) for tropospheric approaches. The MUSICA project (MUlti-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water) within the scope of NDACC aims to characterize long-term observations of the ratio of water vapor isotopologues in the lower/middle troposphere (Schneider et al., 2012, 2016; Barthlott et al., 2017). WV profile retrieval strategies (Schneider et al., 2006; Schneider et al., 2012; Schneider and Hase, 2009; Schneider et al., 2016; Schneider et al., 2017; Barthlott et al., 2017) have been developed and tested. Vogelmann et al. (2015) studied the spatial-temporal variability of WV in the free troposphere (Zugspitze, Germany) by ex-

ploiting the geometry of measurements of differential absorption lidar (DIAL) and FTIR. In particular, they assessed the variability in short scales, i.e., few kilometers and minutes.

In this work, we evaluate the accuracy and precision of WV profiles using a standard retrieval inversion with ground based FTIR measurements. For the first time, the retrieval validation uses coincident and well-characterized balloon-borne in-situ 5 NOAA frost point hygrometer (FPH) measurements (Hall et al., 2016). The FPH measurement technique has been used as reference to assess the accuracy of radiosonde relative humidity measurements due to their high vertical time resolution and low uncertainties (Suortti et al., 2008; Hurst et al., 2011a). ~~Understanding the significance of the WV a priori profiles is important for WV due to its high temporal and spatial variability. Especially, for un-smoothed comparisons with the With the~~ goal to assess WV vertical gradients ~~. In this work, we assess we studied both~~ the influence of ~~the~~ different WV a priori profiles 10 ~~in the retrieval of WV at several altitudes and the smoothing of highly-resolved FPH profiles.~~ Finally, ~~it is well-known that the ubiquitous~~ strong WV absorption signatures interfere in the retrieval of other gases. However, there is a lack of quantitative effect of WV at different altitudes. ~~This study also provides a quantitative assessment of the impacts of WV. A second major part of this work seeks to use FPH profiles as the ground-truth WV and quantitatively assess the impacts of other typical WV profiles~~ in the retrieval of selected tropospheric gases, hydrogen cyanide (HCN), carbon monoxide (CO), and ethane (C₂H₆), 15 using NDACC standard retrievals.

2 Measurements

2.1 Free tropospheric and boundary layer FTIR sites

FTIR direct solar IR absorption spectra are measured under clear-sky conditions in two different locations: (1) Boulder, Colorado (hereafter BLD; 40.40° N, 105.24° W, 1600 m.a.s.l) and (2) Mauna Loa, Hawaii (hereafter MLO; 19.40° N, 155.57° 20 W, 3400 m.a.s.l). The spectra at BLD have been recorded using a Bruker 120 HR spectrometer operated since 2010 following standard measurement protocols of the Infra Red Working Group (IRWG)/NDACC (<http://ndacc.org>). The instrument is located in the foothills laboratory of the National Center for Atmospheric Research (NCAR) situated in the front range of the Rocky Mountains and within the planetary boundary layer. Previous studies have used the BLD dataset for satellite validation 25 of NH₃ (Dammers et al., 2017), mobile low resolution FTIR validation of NH₃ and C₂H₆ (Kille et al., 2017); and analysis of gases emitted by oil and natural gas development (Franco et al., 2016; Tzompa-Sosa et al., 2016). The MLO instrument has 30 been part of the long-term activities of the IRWG/NDACC. First IR solar absorption spectra were recorded at MLO in 1991 using a Bomem DA02. In 1995 a Bruker 120 ~~M-HR~~ started to operate and was upgraded in 2011 to a Bruker 125 HR. The high altitude site at MLO is normally above the boundary layer and the measurements are sensitive mainly to free tropospheric and stratospheric air masses. At both sites the spectra are recorded using optical band pass filters maximizing the signal to noise ratio (SNR) over the near and mid-infra red spectral domain with a nominal spectral resolution of ~~0.0035~~ ~~0.004~~ cm⁻¹ (~~optical path difference of 250 cm~~) using liquid nitrogen-cooled InSb and MCT detectors and a KBr beam-splitter (Hannigan et al., 2009).

2.2 Balloon-borne NOAA Frost Point Hygrometer

Highly precise and accurate in situ measurements of tropospheric and stratospheric WV over Boulder, Colorado, and Hilo, Hawaii, are performed with balloon-borne FPHs by the Global Monitoring Division of NOAA's Earth System Research Laboratory (ESRL). These measurements are also part of the GCOS Reference Upper Air Network (GRUAN) and the NDACC.

5 At both sites, balloon-borne FPHs are launched once per month, preferably during conditions of low winds and clear skies. The Boulder measurements started in 1980 and are launched at Marshall Field Site (1743 m.a.s.l), 10.5 km south of the BLD FTIR measurement site (Oltmans et al., 2000; Scherer et al., 2008; Hurst et al., 2011b). Monthly NOAA FPH soundings at Hilo started in 2010 and the balloons are launched from the National Weather Service facility at Hilo International Airport (10 m.a.s.l), 58.0 km east of MLO. In this paper we emphasize the comparisons at BLD due to the shorter distance between the 10 FTIR and balloon launch site, although we perform identical comparisons and present results from MLO as well.

A thorough description of the FPH measurement technique has been described [elsewhere \(Hurst et al., 2011b; Hall et al., 2016\)](#) in [Hurst et al. \(2011b\)](#) and [Hall et al. \(2016\)](#). Briefly, the basic principle is to condense WV from a stream of air onto a small, gold-plated mirror using a cryogenic liquid to continually cool the mirror. Once a thin condensed layer is deposited on the mirror, pulses of heat are applied as needed to maintain a stable layer of condensate. Changes in frost (ice) coverage are detected 15 by measuring the mirror reflectivity using a small LED-based infrared beam and a photodiode. The amount of heat applied is rapidly adjusting to produce a stable frost layer, at which point the temperature of the mirror (frost point temperature) is a direct measure of the partial pressure of WV in the air stream above it via the Goff-Gratch equation (Goff, 1957). The water vapor mixing ratio is calculated by dividing the WV partial pressure by the dry atmospheric pressure. Since a FPH fundamentally makes temperature measurements, only the thermistor embedded in each mirror requires calibration. Each thermistor is 20 calibrated using NIST traceable standards (see Hall et al. (2016)). A recent detailed analysis of WV mixing ratios measured by the NOAA FPH shows the uncertainties (2-sigma) are < 12 % for the 0 - 5 km altitude layer, < 8 % for 5 - 13 km, and < 6 % for 13 - 28 km (Hall et al., 2016). The NOAA FPH vertical profile data employed here are 0.25 km vertical averages and their standard deviations that are calculated from the measurements made at 5-10 m vertical resolution during balloon ascent.

3 Retrieval of water vapor from FTIR

25 Prior to the retrieval of WV from the solar absorption spectra a quality control of each measurement is carried out, i.e., visual inspection of spectra and assessment of the SNR. As mentioned in Sect. 2.1, we only use spectra taken during cloud-free conditions. The spectra are analyzed using the retrieval code SFIT4 [v4.0.90.9.4](#), which has been improved from its predecessor SFIT2 (Pougatchev et al., 1995; Rinsland et al., 1998; Hase et al., 2004). SFIT4 derives vertical profiles and the corresponding [density vertical columns from the total vertical columns by exploiting](#) pressure broadening and temperature dependency of 30 specific absorption lines. The overall retrieval follows the optimal estimation method applied to several micro-windows. The inverse problem is ill-posed and the solution is constrained by an a priori profile (x_a) and [its](#) covariance matrix (S_a), which ideally should represent the natural variability of the WV profile from climatological records (Rodgers, 2000; Rodgers and Connor, 2003). Section 4.3 describes in more detail the [effect of the a priori and the](#) different a priori profiles used in this

study. In many cases \mathbf{S}_a is not well-known due to lack of long-term highly-resolved profiles and and an ad hoc constraint are is used (e.g., Vigouroux et al. (2015)). The constrained solution is important in order to avoid many different atmospheric states in the minimization of the cost function following a Constraining is important to select the solution which, among the possible solutions of the ill-posed inversion, is the most likely given prior knowledge. The forward model is non-linear and the following Gauss-Newton iteration :-

$$\mathbf{x}_{i+1} = \mathbf{x}_a + \mathbf{S}_a \mathbf{K}_i^T (\mathbf{K}_i \mathbf{S}_a \mathbf{K}_i^T + \mathbf{S}_e)^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}_i) + \mathbf{K}_i (\mathbf{x}_i - \mathbf{x}_a)] \quad (1)$$

where \mathbf{x}_{i+1} is the retrieved state vector for the $(i+1)$ th iteration, \mathbf{K} is the weighting function or Jacobian of the forward model (\mathbf{F}) calculated at each iteration, \mathbf{S}_e is the measurement noise covariance matrix, and \mathbf{y} is the measurement state vector (Rodgers, 2000).

Many of the spectral windows used to retrieve NDACC standard gases contain WV absorption signatures. Retrieved WV is often used in Accurate WV profiles are required for the retrieval of other gases, because accurate quantification of the interfering WV reduces retrieval uncertainty. WV can be retrieved using multiple micro-windows range of absorption features since it absorbs from the near to far infrared wavelengths. With the goal to characterize this 'pre-retrieved' best characterize this WV we use retrieval settings that are commonly used among NDACC sites. We use the 2600 - 2840 cm^{-1} spectral domain region to simultaneously retrieve H_2O and the isotopolog HDO. In this study, we focus only on H_2O . We use spectral micro-windows that are not identical to those of current MUSICA version (Barthlott et al., 2017) and perform the inversion on a linear scale (instead of a logarithmic scale used by MUSICA). A short summary of the four micro-windows and interfering species included in the analysis is given in table 1. These micro-windows have been chosen to maximize the information content and minimize total error. The spectroscopic data used here is based on the line-by-line portion of the HITRAN 2008 (Rothman et al., 2013). The errors in the reported line parameters are described in section 3.1 and are used to estimate the systematic uncertainty in the retrieval. Most of the interfering species are fit as a scaling of the a priori vertical profile (CO_2 , N_2O and HCl) with the exception of CH_4 which is fit as a profile in micro-window two, three, and four. The SNR determines how much influence the spectra has in each micro-window versus \mathbf{S}_a matrix is specified at each layer as a fraction of the a priori ,as well as to characterize the measurement error described in section 3.1. In order to prevent sporadic vertical profile oscillations due to relaxed covariance matrices we implement ad-hoc diagonal elements of \mathbf{S}_a with profile, which allows for a linear scaled retrieval. We adopted a maximum variability of 50 % at the surface % in the diagonal covariance and exponentially decreasing by altitude with a inter-layer thickness correlation coefficient. A Gaussian correlation with a . In order to prevent sporadic vertical profile oscillations, we include a Gaussian correlation length of 25 km is used for in the off-diagonal elements of \mathbf{S}_a . This \mathbf{S}_a has been optimized in order to obtain similar information content for all a priori presented in section 4.3, a requirement for efficient processing of decades of NDACC spectra. The instrumental line shape (ILS) has been fixed with a unity modulation efficiency and ideal no phase error. Schneider et al. (2012) pointed out that the The ILS does not play an important role in the WV error budget and is of lower importance for tropospheric WV retrievals (Schneider et al., 2012).

Table 1. Micro-windows for H₂O retrieval including interfering gases retrieved within those micro-windows. Column gases are those retrieved by profile scaling of initial profile while profile retrieval is done for the profile gases column.

Micro-window [cm ⁻¹]	Profile Gase(s)	Column Gase(s)
1) 2611.40 - 2613.40	HDO	CO ₂
2) 2659.00 - 2661.00	HDO, CH ₄	CO ₂
3) 2819.00 - 2819.80	H ₂ O, CH ₄	N ₂ O, HCl
4) 2829.80 - 2839.40	H ₂ O, CH ₄ , HDO	-

Inputs into SFIT4 include vertical profiles of pressure, temperature, and the volume mixing ratios (VMR) of the atmospheric gases included in the fit. Preceding to the retrieval, SFIT4 employs the Air Mass Computer Program for Atmospheric Transmittance/Radiance Calculation (FSCATM) ray tracing ~~code module~~ to calculate the atmospheric path (Hannigan et al., 2009). The input pressure and temperature vertical profiles are obtained from the National Center for Environmental Prediction (NCEP) reanalysis based on the NCEP/NCAR analysis/forecast system to perform data assimilation using past data from 1948 to the present ([Kalnay et al., 1996](#)) ([Finger et al., 1993](#); [Wild et al., 1995](#); [Kalnay et al., 1996](#)). These profiles are obtained directly from NDACC (<http://ndacc.org>). These are daily average profiles that extend to up to 0.4 mb (approximately 50 km). Above 0.4 mb we use monthly mean pressure and temperature profile from an average of a 40 year simulation (1980-2020) of the Whole Atmosphere Community Climate Model (WACCM) (Garcia et al., 2007). These profiles are merged using a cubic spline interpolation for pressure and a quadratic spline interpolation for temperature.

We examined the effect of using more temporally refined temperature profiles. In general, the ~~root mean square error (rmse) of the fit between the six hourly data of six hourly temperature profile from the ERA-I reanalysis model~~ produced by the European Center for Medium-Range Weather Forecasts (ECMWF) (Dee et al., 2011) ~~and daily~~, follows the daily average temperature profile shape very well for both sites. The root mean square error (rmse) between the six hourly data of ERA-I and daily average temperature is less than 0.5% ~~of the mean temperature~~ using 2013 data ~~and the bias of the fits are insignificant indicating that the daily mean temperature profile is adequate for the retrievals for both BLD and MLO and the biases are less than 0.25 % for BLD and less than 0.1 % for MLO. These results suggest daily mean temperature should be adequate for retrievals but we further investigated the sensitivity of water vapor to this variability and found that water vapor agrees within 1 % if using the daily average profile. The temperature profile uncertainty is considered in the error analysis in section 3.1.~~ With the exception of WV (see section 4.3), VMR input mean profiles of all other gases are taken from the ~~mean of a 40 year run of WACCM~~.

3.1 Characterization and error budget

The mean retrieval fit of the four micro-windows between 2010-2016 at BLD is shown in Fig. 1. The small systematic residual structures (black lines) are likely caused by spectroscopic parameter error but in general the magnitude of residuals is low and within noise level (< 0.1 %).

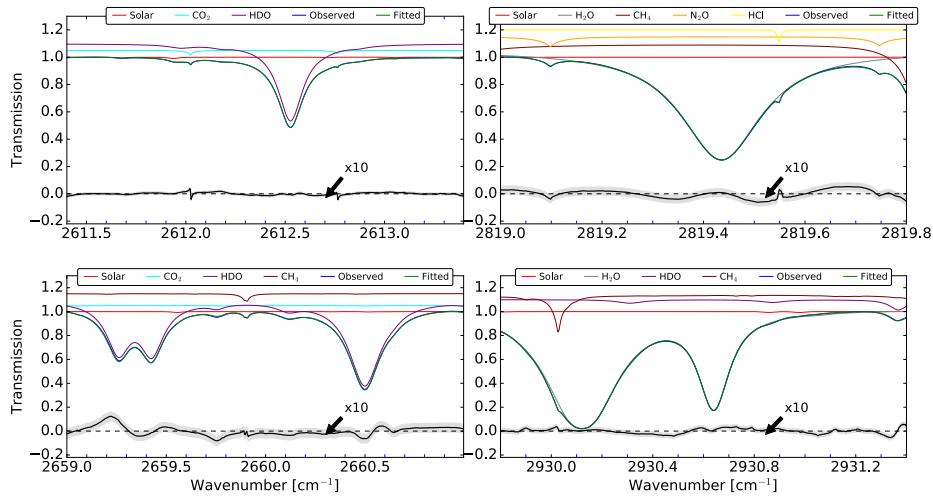


Figure 1. Mean retrieval fit between 2010 - 2016 for the spectral intervals of WV. The observed and fitted lines are blue and green respectively. The absorption contribution for the different species are also shown in each micro-window. The bottom black lines represent the mean residual and the gray shadow are the standard deviation. Note that for visibility the residuals have been multiplied by 10.

The information content of the retrieved WV vertical profile is characterized ~~by means of within~~ the averaging kernel matrix, **A**:

$$\mathbf{A} = (\mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K} \quad (2)$$

The rows of the mean **A**, known as averaging kernels (AK) ~~at BLD~~ obtained between 2010 - 2016 ~~and~~ color coded by 5 altitude below 20 km are shown in Fig. 2a ~~for BLD~~. The maximum values are located at the surface, then they decrease and remain steady to about 8 km and eventually decrease to zero above 12 km. This indicates that most of the information content ~~obtained from WV pressure dependence of the absorption features is limited to the is derived from the lower~~ troposphere. The mean total column averaging kernel (TAK) is shown in Fig. 2b. ~~Typically~~ ~~Typically~~, a unity TAK indicates that the retrieval is not biased, while values of the TAK lower than unity indicate underestimation and larger values than unity indicate overestimation 10 ~~of the real WV with respect to the a priori state vector~~. Hence, below 3 km the retrieval may underestimate, between 3-8 km overestimate, and between 8-12 km underestimate the real WV magnitude. The mean number of ~~degrees of freedom (DOFs)~~ ~~DOFs~~, given by the trace of the **A**, are 2.4 and indicate the total number of independent pieces of information in the retrieval. The vertical profile of the cumulative sum of DOF is shown in Figure 2c and shows that the first DOF is given in 15 the layers below 3 km, the second DOF is given between 3 - 6 km, and the rest above. Further optimization of the retrieval strategy might improve the **A** but as explained before ~~one of the goals is to assess the current retrieval strategy, therefore we do not investigate retrieval constraints further. At MLO the vertical sensitivity is similar but starting at 3.5 km. A proper comparison between FTIR and in-situ sonde profiles would require to smooth the in-situ measurements using the FTIR AK and a priori to account for lower resolution measurement's smoothing (Rodgers and Connor, 2003). However, as pointed out by Schneider et al. (2006) the information of the WV **A** is limited due to its high variability through the troposphere. Additionally,~~

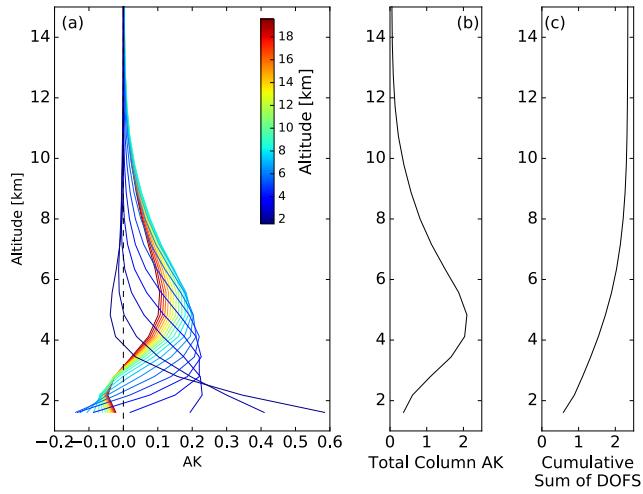


Figure 2. (a) FTIR mean row averaging kernels; (b) mean total column averaging kernel; and (c) cumulative sum of DOF of WV obtained in BLD from 2010 - 2016.

~~smoothing the in-situ measurements would require using the a priori profile, which in turn may be highly biased to the real atmospheric state. A goal of the present study is to determine whether the highly structural variability of WV can be retrieved with FTIR measurements, hence the comparison with in-situ sonde measurements are carried out without a smoothing.~~

5 SFIT4 estimates an uncertainty budget that combines random, systematic, and smoothing sources following the formalism given in Rodgers (2000). The most important random error is normally the retrieval noise characterized with the SNR in the spectral region of interest. The error covariance matrix (\mathbf{S}_n) is calculated with the following equation:

$$\mathbf{S}_n = \mathbf{G}_y \mathbf{S}_e \mathbf{G}_y^T \quad (3)$$

where the gain matrix \mathbf{G}_y represents the sensitivity of the retrieval to the measurement and is related with the averaging kernel as $\mathbf{A} = \mathbf{G}_y \mathbf{K}$. Currently, the diagonals of the \mathbf{S}_e matrix are constructed using the square of the inverse of the SNR obtained 10 from the noise in the spectra of interest, and off diagonal elements are not considered. The retrieval of WV is actually an estimate of a state smoothed by the averaging kernel. The difference between these two states is given by the smoothing error (\mathbf{S}_s):

$$\mathbf{S}_s = (\mathbf{I} - \mathbf{A}) \mathbf{S}_a (\mathbf{I} - \mathbf{A})^T \quad (4)$$

15 where \mathbf{I} is a unit matrix. The smoothing error is treated separately and not included in the total error analysis because \mathbf{S}_a is normally not well known and consequently is often simplified. The model parameter error represent the errors in the forward model parameters such as temperature, solar zenith angle (SZA), and spectroscopic parameters. These errors can contain both systematic and random components. We obtain the model parameter covariance matrix as:

$$\mathbf{S}_b = (\mathbf{G}_y \mathbf{K}_b) \mathbf{S}_b (\mathbf{G}_y \mathbf{K}_b)^T \quad (5)$$

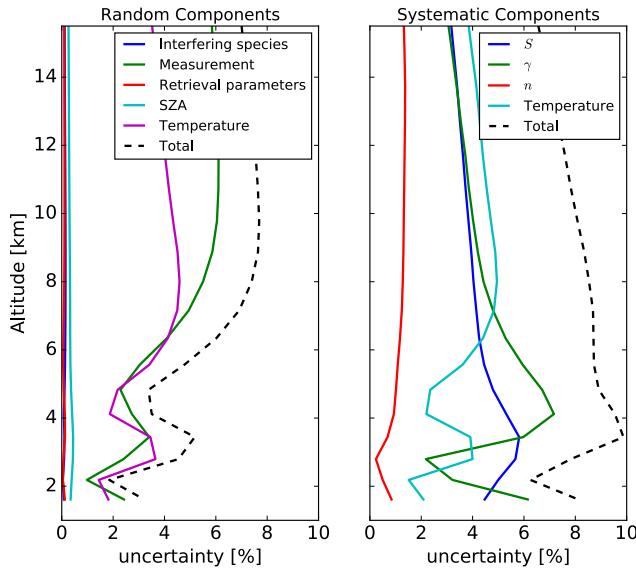


Figure 3. Mean vertical profiles of the most important random (left) and systematic (right) uncertainty components for the retrieval of WV in BLD from 2010 - 2016.

Where, \mathbf{S}_b is the error covariance matrix on the and \mathbf{K}_b the weighting function matrices of the forward model parameters. The largest contributors are considered here and are the absorption line parameters, temperature profiles, and SZA. The uncertainty of the absorption line parameters, i.e., line intensity (S), air-broadened half width (γ), and temperature dependence of γ (n), are taken from the lower limit reported in HITRAN 2008 (Rothman et al., 2013). These uncertainties are only considered systematic and the errors reported in HITRAN for WV are 5, 1, and 10 % for S , γ , and n , respectively. Furthermore, uncertainty due to the retrieved interfering species are also considered. The error in the temperature profile is considered to have both systematic and random components.

These errors have been quantified with the mean (systematic) and standard deviation (random) of the difference of long-term comparisons between NCEP profiles with radiosondes launched near the sites and/or ERA-I reanalysis. The measurement noise error is estimated with the square of the inverse of the SNR as diagonal elements in the covariance matrix. The pointing accuracy in the SZA is considered random and has been characterized with an error of 0.15° .

Figure 3 shows the random and systematic vertical profile uncertainties in percent with respect to the mean mixing ratio. The major systematic components in the lower troposphere are the absorption line parameters S and γ but in the upper troposphere the temperatures contributes equally. The temperature and measurement noise are the main components of the random uncertainty. The final uncertainty is estimated from the error propagation of all components and is lower than 10 % below 4 km and about 10 % above. The instrumental line shape uncertainty plays a minor role in the total error budget.

4 Comparison of water vapor vertical profiles

The total number of sonde observations are 90 at Boulder and 70 at Hilo from 2010 to 2016. The overall number of coincident dates of measurements under ideal conditions are 56 and 36 for BLD and MLO, respectively. Figure 4 presents a rough qualitative comparison of selected WV profiles obtained with NOAA FPH measurements and FTIR retrievals in BLD.

5 ~~The daily mean ERA-I (henceforth ERA-d) a priori profiles used in the retrievals are also shown.~~ To retain high vertical variability the FPH profiles are shown in 0.25 km vertical averages of the sonde's ascent measurements ([black continuous lines](#)). The FTIR profiles ([in blue](#)) represent the average profile weighted by the error and the blue shading depicts the uncertainties propagated using the individual profiles within 2 h of the FPH launch. The ~~daily mean ERA-I (henceforth ERA-d) a priori profiles used in the retrievals are also shown in gray.~~

10 ~~To quantitatively compare both measurements the high vertical resolution balloon-borne profiles are re-gridded onto the altitude grid of the FTIR retrieval by means of a linear interpolation. For BLD the nearest FPH point to the surface is typically few hundred meters above the first grid point of the FTIR. In this case, we assume homogeneous WV close to the surface and use the nearest-neighbor point. A proper comparison between FTIR and in-situ sonde profiles requires smoothing the in-situ measurements using the FTIR AKs and a priori profiles to account for its lower vertical resolution capability~~

15 ~~(see equation 4 in Rodgers and Connor (2003)). Red profiles in Fig. 4 represent smoothed FPH profiles. As pointed out by Schneider et al. (2006) the information of the WV AK is limited due to its high variability through the troposphere. A goal of the present study is to determine the extent of vertical structure gradients of retrieved WV profiles, hence the comparison with in-situ sonde measurements are carried out mainly without smoothing. However, results are also presented for smoothed comparisons following the formalism given in Rodgers and Connor (2003).~~

20 ~~The temporal variability and its effect are studied in section 4.1. The To some extent the retrieved WV profiles capture the vertical structure gradients identified with the in-situ NOAA FPH even though the a priori profile may be biased and smooth (see for example 2010-09-14 and 2010-11-05). For comparison, Fig. S1 in the Supplement also shows the same figure but smoothing the sonde profiles using the FTIR averaging kernel and a priori to account for lower resolution measurement's smoothing (Rodgers and Connors, 2003). However, it is clear that smoothing diminishes the real variability, which is actually captured by the retrieval.~~ Figure 5 shows the same but for selected vertical profiles at MLO. The near-surface mixing ratios at this high-altitude site are significantly lower and the profiles show steeper vertical gradients than at BLD. Note that the FTIR (MLO) and FPH (Hilo) are about 60 km apart and might have sampled different air masses. In BLD the launch [site](#) of the FPH is only 10 km south of the ground-based FTIR. ~~The generalized comparison using smoothed FPH profiles at MLO are also shown in the Supplement (see Fig. S2).~~

25 ~~Comparisons of WV vertical profiles for selected dates obtained with in-situ NOAA FPH measurements (black) and FTIR retrievals (blue) in BLD. The ERA-d WV used as a priori is shown in gray. The dates are shown at the top of each plot. The FTIR profiles represent weighted mean profiles using retrievals within two hours of the radiosonde launch. The blue filled shadow area represents the standard error propagation using the uncertainty in individual retrievals. The gray shaded are of~~

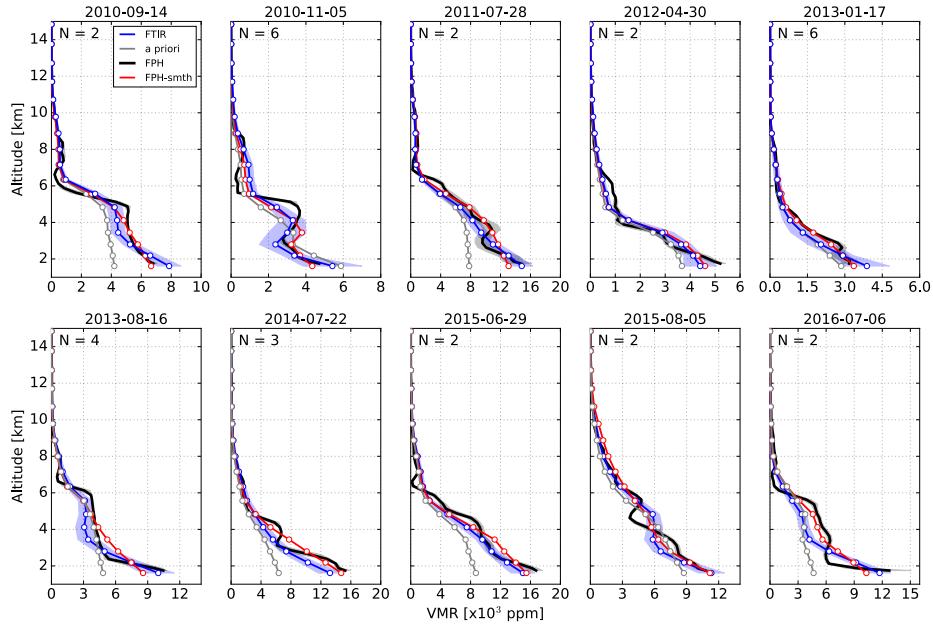


Figure 4. WV vertical profiles for selected dates obtained with un-smoothed in-situ NOAA FPH measurements (black) and FTIR retrievals (blue) in BLD. The ERA-d WV used as a priori is shown in gray. The dates are shown at the top of each plot. The FTIR profiles represent weighted mean profiles using retrievals within two hours of the radiosonde launch. The blue filled shadow area represents the standard error propagation using the uncertainty in individual retrievals. The gray shaded area of FPH profiles are the 1-sigma standard deviation of each mixing ratio. The number of retrieved profiles within 2 hours is shown on the upper-left of each panel.

FPH profiles are the 1-sigma standard deviation of each mixing ratio. The number of retrieved profiles within 2 hours is shown on the upper-left of each panel.

Same as Fig. but for MLO.

To quantitatively compare both measurements the high vertical resolution balloon-borne profiles are re-gridded onto the 5 altitude grid of the FTIR retrieval by means of a linear interpolation. For BLD the nearest FPH point to the surface is typically few hundred meters above the first grid point of the FTIR (see Fig. 4). In this case, we assume homogeneous WV close to the surface and use the nearest neighbor point. Due to the limited number of DOF we do not aim to compare every grid point but combine grid points to assess several layers maximizing the number of points and characterizing yet characterizing the boundary layer, free troposphere, and upper troposphere - lower stratosphere. The following layers have been chosen for BLD: 10 1.5-3.0, 3.0-5.0, 5.0-7.5, 7.5-10, 10-13, and 13-17 km above sea level (asl) and for MLO: 3-5.5, 5.5-7.5, 7.5-10, 10-13, 13-16, and 16-20 km asl. These layers have been chosen so that they include three standard IRWG FTIR grid points. Comparison of ground-based remote sensing with balloon-borne in-situ measurements is challenging due to spatial-temporal variability. The temporal and spatial variability are characterized in the next two sections followed by the quantitative comparison between FTIR and NOAA FPH.

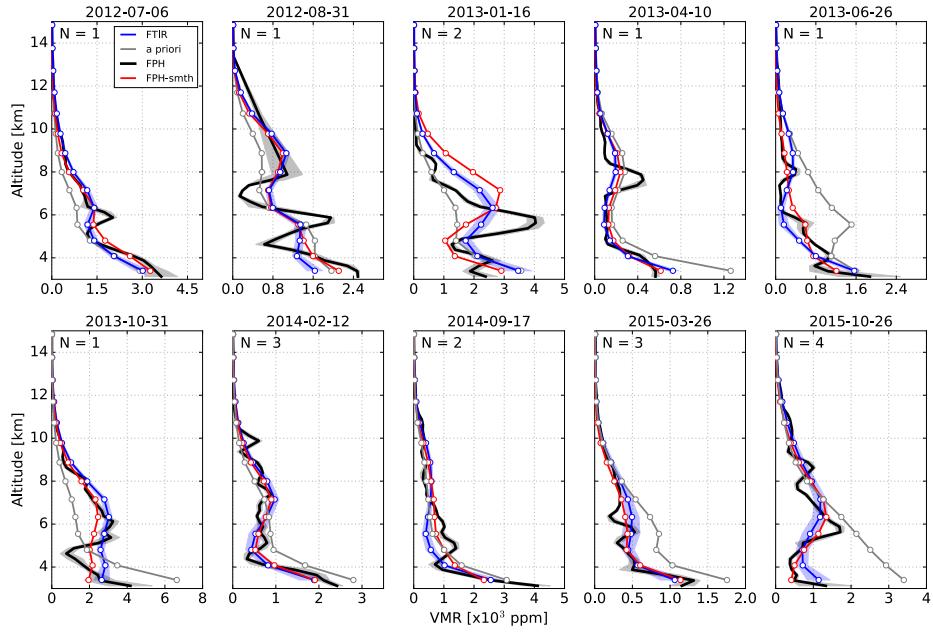


Figure 5. Same as Fig. 4 but for MLO.

4.1 Temporal variability

Due to the lack of independent time-resolved WV vertical profiles we use daily FTIR observations to assess the temporal variability. Figure 6 shows the number of dates and profiles and the variability of WV in percent for several layers as function of the length of time interval starting from 0 to 3 minutes and gradually increasing, e.g., 0 to 10, 0 to 30, 0 to 60 minutes, 5 etc. The retrievals produced during these time intervals are used to calculate the temporal variability using the ratio of the standard deviation to the mean values at several altitude layers. This approach is sensitive only to the variability observed by the FTIR, however the real variability might be greater because of potential lost variability during retrieval smoothing. This proxy for variability has been estimated using dates during coincident measurements between sondes and FTIR. The number of dates and profiles are roughly the same below 10 min, indicating the time that the FTIR takes to start a new measurement 10 using the same band pass filter for a standard set of observations. The number of profiles starts to increase from the number of dates after 15 min. The variability in BLD among different layers does not vary substantially and they remain within 1 - 2 % of each other, indicating similar relative variability within all the different tropospheric and stratospheric layers. In BLD the variability starts to increase from about 1 % in 30 min to 6 % in 240 min. In contrast, at MLO the variability is different among 15 layers. A variability up to 9 % is found for the layer close to the instrument altitude (3 - 5.5 km), however the variability is below 5 % for the layer between 5.5 - 7.5 km, and even below 5 about 3 % for the 16-13 - 20-16 km indicating a vigorous fluctuations and strong convection near the MLO site. In general, these findings suggests that the coincidence time interval to avoid variability larger than 2 % is 30 min at BLD and 60 min at MLO. The air mass probed by the FTIR is changing during the

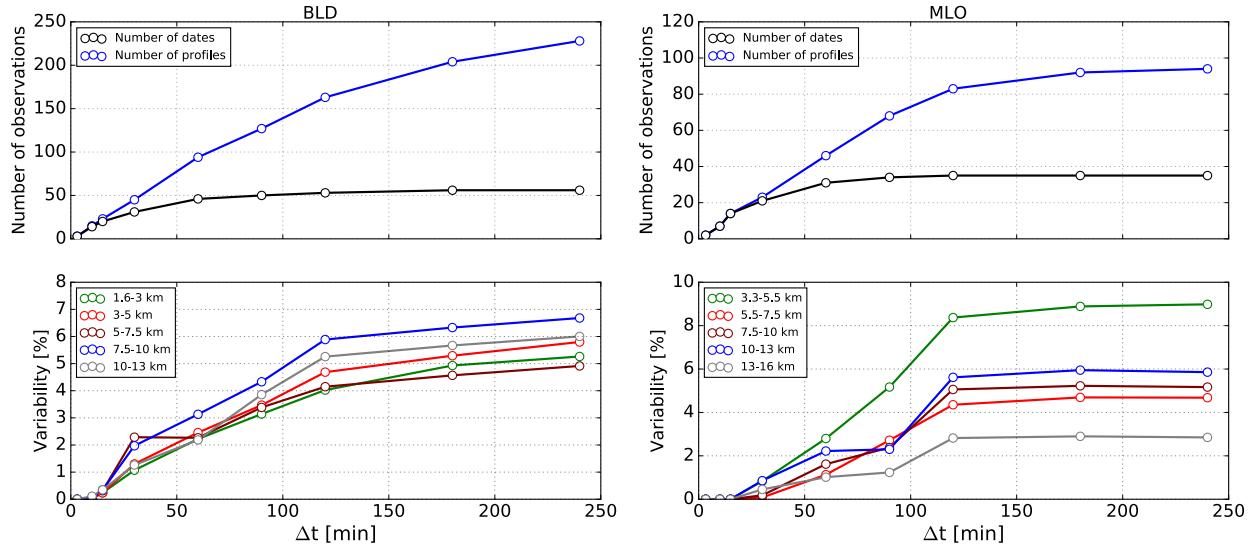


Figure 6. The top panels show the number of dates (black) and profiles (blue) measured by the FTIR at BLD (left) and MLO (right) as a function of the length of the time interval in minutes. The bottom panels show the temporal variability in percent estimated with the ratio of the standard deviation to the mean values for several layers as a function of the length of the time interval. The length of the time intervals are defined as increasing temporal window, e.g., 0 - 30, 0 - 60, 0 - 120 minutes, and the number of retrievals in each window is used to calculate the variability.

day due to the line of sight to the sun moving constantly such that after some time the spatial variability may play an important role. Vogelmann et al. (2015) estimated that the spatial mismatch may play a role for intervals longer than 30 min. The spatial mismatch is described in the next section.

4.2 Spatial mismatch

5 If the spatial mismatch between the FTIR and sonde is considerably large they might probe distinctive air masses. Hence, natural WV variability would affect a meaningful comparison (Sussmann et al., 2009; Vogelmann et al., 2015). A thorough assessment of the spatial variability of WV error component due to spatial difference between the sonde and FTIR would require measurements of an extensive area simultaneously and at different altitudes. In order However, this is hard to derive due to lack of such observations. In this section, we aim to estimate the spatial mismatch we between the sonde location at

10 various altitudes and FTIR maximum sensitivity. We calculate the horizontal distance between the sonde location and the line of sight of the FTIR. The effective horizontal position sensitivity of the FTIR depends on the sun-pointing geometry and the vertical WV profile distribution. We adopted a methodology applied by Vogelmann et al. (2015) to estimate this effective horizontal position. This method assumes that the FTIR sensitivity is located at the point where the viewing direction of the instrument meets the altitude level of the mass weighted WV profile. Using the mass weighted WV of all sonde profiles we

15 estimate roughly an altitude of 3.8 ± 0.9 km in BLD. Using this altitude and the SZA the horizontal distance from the ground-

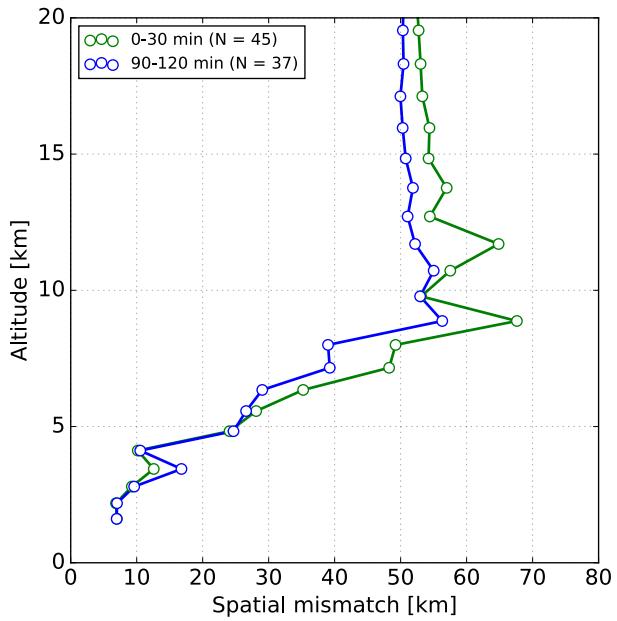


Figure 7. Vertical profile of the horizontal spatial mismatch between FTIR and sonde profiles in BLD. As an example two coincident time intervals are used.

based site is calculated for every measurement. Then, using the solar azimuth angle the latitude and longitude are calculated after having traveled the given distance on the given bearing. Once the location is found the haversine formula is applied to determine the great-circle distance between two locations (Korn and Korn, 2000). At BLD the mean distance with respect to the FTIR site location is 6.0 ± 4.0 km south making the initial spatial mismatch with the sonde launch about 6.5 km. At MLO 5 the mass weighted WV profile is 6.0 ± 0.6 km and the initial spatial horizontal mismatch is 47.0 km (see Fig. S3-S1 in the Supplement). Consequently, even co-located sonde launches may not exactly probe the same air mass.

The spatial mismatch at different altitudes depends on the sonde trajectory and the location of the FTIR sensitivity. At BLD the GPS location at every sonde altitude is available for almost all profiles and the horizontal distances between all altitudes and the FTIR sensitivity on the earth are calculated. Figure 7 shows the mean spatial mismatch between the FTIR and the sonde 10 profiles for the coincident time intervals of 0 - 30 and 90 - 120 minutes. As mentioned above, the initial spatial difference close to the surface is about 6 km. For the 0 - 30 min interval the horizontal difference is below 10 km below 4.5 km altitude, similarly for the 90 - 120 min, except for one altitude, which is greater than 15 km. Above 5 km altitude the spatial mismatch starts to increase. A rapid significant increase in the spatial mismatch is identified above 5 km for both 0 - 30 and 90 - 120 min coincident time intervals. Interestingly, the greatest horizontal difference is found for the 0 - 30 min interval with maximum 15 values of about 70 km. This analysis shows that the spatial mismatch depends on the complex convective dynamics and not only in the coincidence time interval. Nevertheless, only short temporal coincidence differences are encouraged to avoid temporal WV fluctuations as shown above.

4.3 Influence of a priori profiles

The optimal estimation method is influenced by the a priori profile influences because it may bias the solution of equation 1 in two ways, one in the first order weighting and also if the retrieval is not in a linear regime (Kulawik et al., 2008). Since WV is highly variable, even in time scale of hours, using the most accurate a priori might improve the retrieval results. Especially,

5 if the comparison are carried out without smoothed in-situ profiles with aim to capture vertical gradients and to avoid limited information in the averaging kernel. In general, the retrieval of WV can be seeing as an update of the a priori information. In order to study the effect of the apriori, four different a priori profiles are used to retrieve WV, which then are compared with balloon-borne NOAA FPH measurements: (1) a 40 year simulation (1980-2020) of the WACCM monthly mean profiles. WACCM is a global model with 66 vertical levels from the ground to approximately 140 km geometric height, the horizontal 10 resolution is 1.9/2.5° (latitude/longitude) and is part of the NCAR Community Earth System Model (for further details see Garcia et al. (2007); Marsh et al. (2013); Kinnison et al. (2007)) ; (2) daily varying (ERA-d); (3) 6 hourly varying WV vertical profiles (00, 06, 12, and 18 UTC) obtained from ERA-I (ERA-6). In this case, the closest in time to the measurements is used. ERA-I profiles extend to 1 mb and then are merged with WACCM monthly mean profiles of WV using a spline interpolation. We take the closest ERA-I grid point to represent the a priori at each station; and (4) daily varying NCEP/NCAR (NCEP-d) 15 reanalysis WV profiles (Kalnay et al., 1996). Since the spatial resolution of NCEP is lower (2.5 x 2.5°) we interpolate WV spatially to obtain the best WV profile.

We have chosen the above four a priori profiles since they are readily available and commonly used. With the aim being to capture vertical gradients the comparisons are carried out with un-smoothed and smoothed in-situ profiles.

An optimization of the data set is carried out before the quantitative comparison assessment of vertical profiles. The difference 20 between WV retrievals and sonde profiles ($\Delta x = x_r - x_s$) shows a normal distribution centered around zero for the layers defined in section 4. Fig. S4-S2 in the Supplement shows an example of the Δx distribution using ERA-d for the different layers. Extreme outliers are identified for each distribution using the 95th percentile and values above that are filtered out in order to avoid skewed results. Figure S5-S3 shows the 95th percentile of the Δx as a function of the different a priori sources and for different layers. The lowest values are found for both ERA-d and ERA-6, and about 25 % larger values are found for 25 both NCEP and WACCM. Additionally, the difference between WV retrievals and a priori profiles ($x_r - x_a$) provides further evidence in the measured signal and to some extent the variability prescribed by the a priori (Rodgers and Connor, 2003). For example, this difference is about 11 ± 38 % using ERA-6 while for WACCM is about 29 ± 32 % for the first layer. As we expected, from these observations it can be seen that the 40 year WACCM climatology as a priori results in greater deviations compared to ERA-6.

30 A quantitative impact of the different a priori in the retrieval of WV vertical profiles is characterized by means of linear regression and statistical analyses using the layers defined earlier. Since both NOAA FPH and FTIR have uncertainties associated at each altitude we adopted a weighted orthogonal distance regression (ODR) analysis. For a thorough description in weighted ODR applied in atmospheric sciences see Wu and Yu (2018). In order to avoid temporal variability larger than 2 % according to conclusions in section 4.1 a mean WV profile (\bar{x}_r) is obtained within a coincidence time interval of 0 - 30

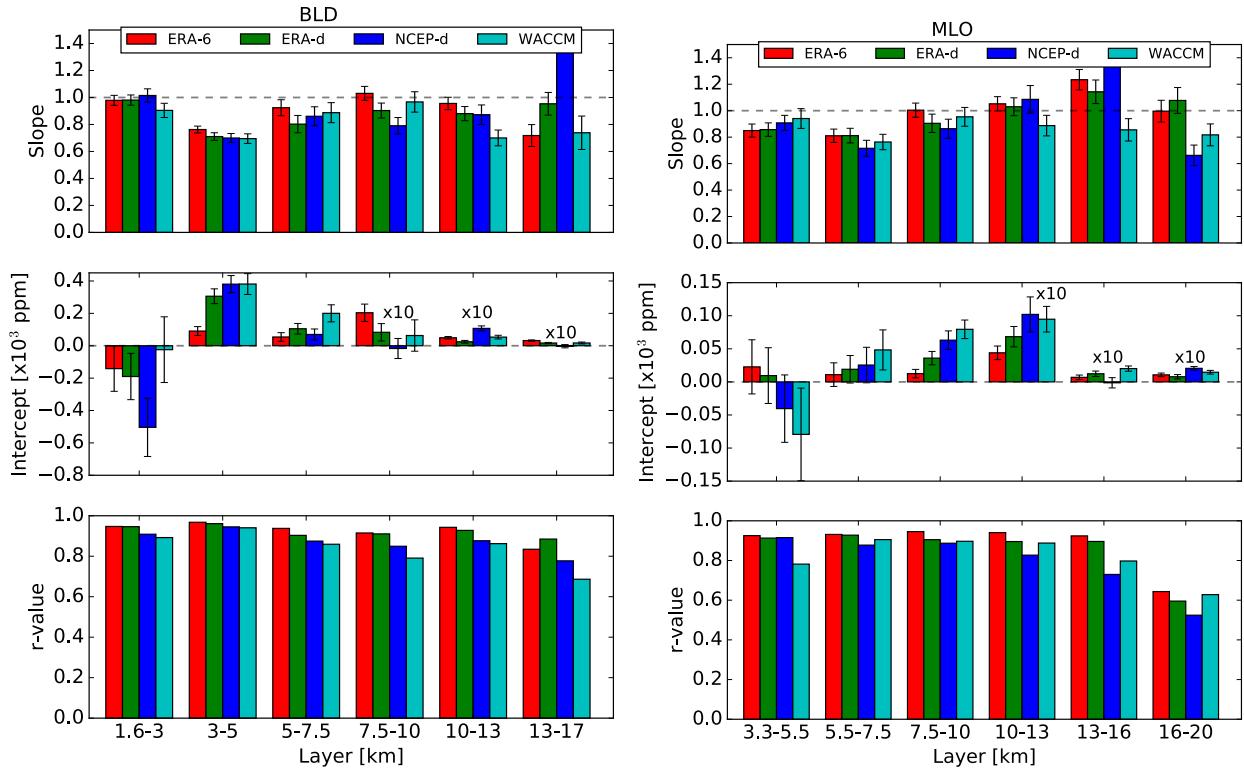


Figure 8. Results of the ODR analysis between the NOAA FPH and FTIR using different a priori profiles at different altitude layers. Error bars represent the standard errors of the estimated parameters. Note that for visibility the intercept obtained in the upper three layers has been multiplied by a factor of 10.

min at BLD and 0 - 60 min for MLO. The NOAA FPH WV mixing ~~rations~~ ~~ratios~~ are used in the abscissa axis and the ODR accounts for uncertainties in both set of measurements. In this case we use the standard deviation of the NOAA FPH and FTIR uncertainty propagated using the individual profiles within the coincident time interval. The final number of vertical profiles used in the comparison are 31 and 30 in BLD and MLO, respectively. Figure 8 shows the slope, intercept, and correlation 5 coefficient (r-value) obtained with the comparison of retrievals using each of the a ~~prioris~~ ~~with the priori with the un-smoothed~~ NOAA FPH at different layers in both sites. The error bars in the estimated parameters are the standard errors. For layers below 10 km the best results are seen with both ERA-I a ~~prioris~~ ~~priori~~. In particular, we found that ERA-6 yields the best comparison 10 with a slope close to unity, the lowest intercept, and a correlation coefficient of 0.95 for the layer of 1.5 - 3 km in BLD. For both sites, the second layer, i.e., 3 - 5 and 5.5 - 7.5 km for BLD and MLO respectively, shows lower slopes likely due to gradients between the top planetary boundary layer and free troposphere that are not captured by the retrievals due to coarse vertical resolution and lower sensitivity (e.g., see Figs. 4 and 5).

For each coincidence profile the bias is characterized with the sum of differences between \bar{x}_r and the sonde (x_s) profiles divided by the number of points (N) in each layer. As described before the number of points in each layer is three. This

definition indicates whether the retrievals under - or overestimate the sonde values. The precision is calculated as $2 \times \sigma / \sqrt{N}$, where σ is the standard deviation. The bar plot in Fig. 9 shows the median bias and precision in ppm and percentage with respect to the mean values of the NOAA FPH for the different layers and a ~~priori~~priori. The error bars in the bias are estimated using the ± 1 -standard error of the distribution. The bias shows ~~a high~~the dependency on the ~~a~~ priori. At both sites the first 5 two layers show negative bias for all a ~~priori~~priori. At BLD the smallest bias is found for the 1.5 - 3 km layer with $-0.001 \pm 0.105 \times 10^3$ ppm ($-0.02 \pm 1.86\%$) for ERA-6 and the highest bias of $-0.27 \pm 0.11 \times 10^3$ ppm ($-4.82 \pm 1.94\%$) for WACCM climatology. The layer between 3 - 5 km shows negative bias between 5.56 % and 11.14 %. Interestingly, NCEP-d yields less 10 biased results in this layer. The layer of 13 - 17 km shows significantly larger values for almost all a ~~priori~~priori ($> 15\%$). The precision does not change significantly among different a ~~priori~~priori. The best precision result in percentage is below 5 % found in the lowest layer of 1.5 - 3 km and the highest values of up to 15 % for layers between 5 - 10 km. As expected based 15 in the ODR analysis higher biases are found at MLO. A Negative bias of about 5 % is found for the 3.5 - 5 km layer, and about 10 % for the 5.5 - 7.5 layer and positive 5 % for the 7.5 - 10 km layer. Surprisingly in both sites WACCM yields lower bias for the layers above 13 km. In general, among all layers the lowest bias are found using ERA-6 and ERA-d for both sites.

The approach described above has been applied in the comparison of FTIR with smoothed FPH profiles. Table 2 presents 15 a summary of the ODR and statistical analysis using ERA-6 for un-smoothed and smoothed FPH profiles at BLD where the spatial mismatch is known and the launch of the sonde is in close proximity to the FTIR location. Among all layers the ODR analysis shows similar results between un-smoothed and smoothed FPH comparisons, however biases are significantly lower for un-smoothed comparisons, indicating the limitation of the AK WV.

5 Influence of WV on gas profile retrievals

20 Absorption of WV is normally ~~expected~~present in the analysis of gases using FTIR measurements. Even optimized micro-windows of gases include the WV and/or isotopologues absorption lines in order to minimize its interference. In this context, WV profiles are included in the retrieval process, ~~a vertical profile is fitted normally for a target gas and other species can also be fitted as profile or simply by single scaling of their a priori profile. This a priori or reference profile may play an important role, especially if it changes significantly diurnally and seasonally. In the case of WV sometimes it~~ of other atmospheric gases. 25 Usually, the most accurate WV profile is recommended. However, highly accurate and co-located WV profile measurements are rare and typically reanalysis based or pre-retrieved WV profiles are used as reference in the retrieval of other gases. In the latest case, WV is retrieved in dedicated micro-windows and then the retrieved WV profile is used in the retrieval of other gases. ~~Normally, WV is again fitted, but now with only one scaling parameter (Vigouroux et al., 2012). So far, however, there~~ (Vigouroux et al., 2012; García et al., 2014; Sepúlveda et al., 2014). Sussmann and Borsdorff (2007) studied the impact 30 of WV interference in the retrieval of carbon monoxide (CO) and further apply a joint retrieval strategy to remove interference errors. There is little published data on the quantitative impact of the WV ~~a priori profile~~ profile using independent co-located WV profiles. Findings from previous sections provide important insights into how well the retrieved WV, and other WV priors, compare with the real WV profile, in this case the NOAA FPH. In this section we investigate the influence of the different WV

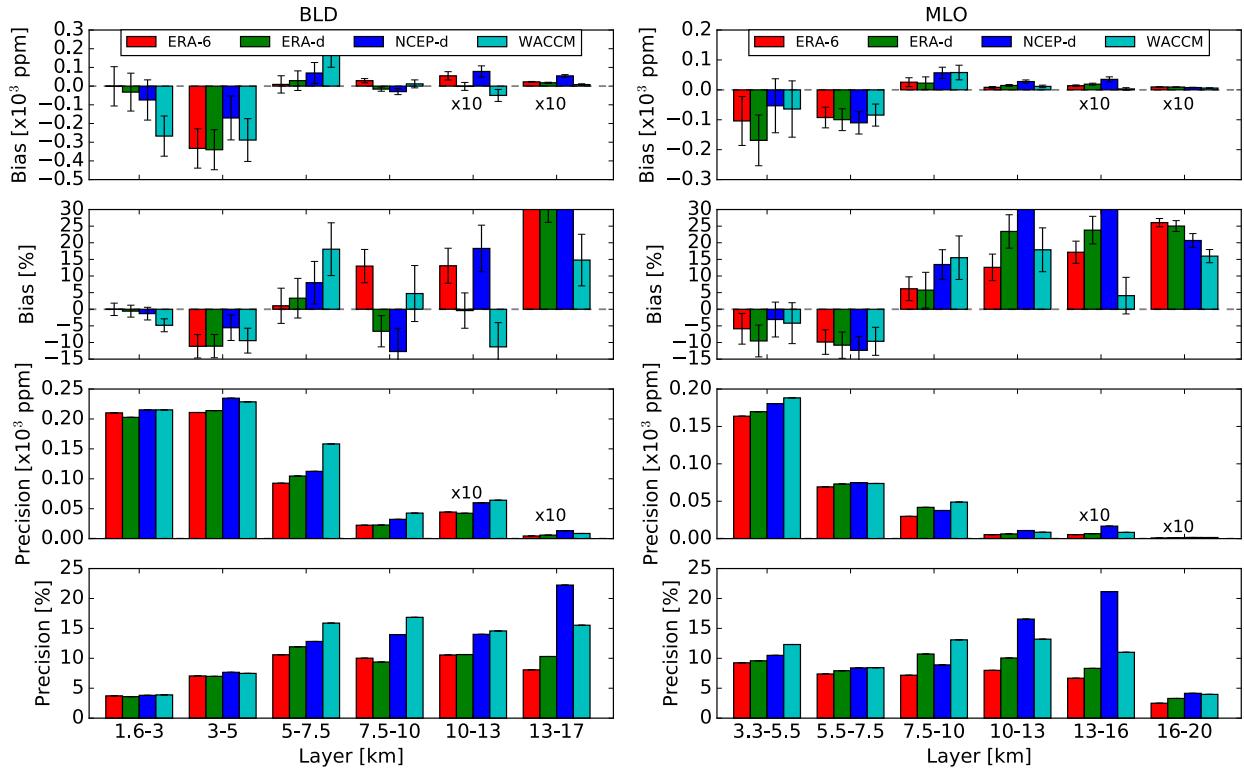


Figure 9. Statistical analysis results (bias and precision) of the FTIR WV retrieved at different altitudes and using different a priori profiles for BLD (left) and MLO (right). Bias and precision are given in mixing ratios and percent with respect to the mean values at each layer. The error bars in the bias represent the standard error of the distribution. Note that for visibility the bias and precision in mixing ratio from the two upper layers have been multiplied by a factor of 10.

sources, we further exploit the FPH measurements in order to study the influence of different and typical WV sources in the retrieval of selected tropospheric gases, i.e., hydrogen cyanide (HCN), carbon monoxide (CO), and ethane (C_2H_6). The WV sources tested are ERA-6, ERA-d, NCEP, WACCM, and retrieved WV profiles, in the retrieval of selected gases. Note that we do not aim to study retrieval strategies of gases or the validation of profile retrievals but rather to show the relative difference with respect to the 'truth' WV profile, in this case the NOAA-FPH higher precision WV profile (FPH measurements). Table 3 presents the three target gases (HCN, CO, and C_2H_6) and a summary of the retrieval settings interfering species with strong and/or weak absorption signatures within each micro-window for all target gases. In all cases, the selected settings have been chosen in order to maximize the information content and minimize the total error in the retrieval. The settings we follow are IRWG/NDACC standard operational retrieval parameters with respect to micro-windows and interfering species. The WACCM climatology is used for a priori profiles of interfering species. Spectroscopic line parameters are adopted from HITRAN 2008 (Rothman et al., 2009, 2013). For the retrieval of HCN we followed a similar approach applied in Paton-Walsh et al. (2010); Vigouroux et al. (2012); Viatte et al. (2014). The settings applied in the CO retrieval are part of an ongoing project in the IRWG

Table 2. Summary of the ODR and statistical analysis using ERA-6 at BLD. Results for un-smoothed (upper level) and smoothed (lower level) FPH comparisons are shown.

Layer [km]	slope	Intercept [$\times 10^3$ ppm]	r-value	Bias [%]	Precision [%]
Un-smoothed					
1.6 - 3.0	0.98 ± 0.04	-0.14 ± 0.14	0.95	-0.02 ± 1.9	3.7
3.0 - 5.0	0.76 ± 0.03	0.09 ± 0.03	0.97	-11.1 ± 3.5	7.0
5.0 - 7.5	0.92 ± 0.06	0.05 ± 0.03	0.94	1.0 ± 5.3	10.6
7.5 - 10.0	1.03 ± 0.05	0.02 ± 0.005	0.91	13.0 ± 5.0	10.0
10.5 - 13.0	0.96 ± 0.05	0.005 ± 0.001	0.94	13.1 ± 5.3	10.6
13.0 - 17.0	0.72 ± 0.08	0.003 ± 0.001	0.83	41.6 ± 4.0	8.1
Smoothed					
<u>1.6 - 3.0</u>	<u>0.95 ± 0.04</u>	<u>-0.05 ± 0.15</u>	<u>0.95</u>	<u>-1.8 ± 1.8</u>	<u>3.6</u>
<u>3.0 - 5.0</u>	<u>0.82 ± 0.02</u>	<u>0.04 ± 0.03</u>	<u>0.97</u>	<u>-11.5 ± 3.0</u>	<u>6.0</u>
<u>5.0 - 7.5</u>	<u>1.0 ± 0.04</u>	<u>0.02 ± 0.02</u>	<u>0.97</u>	<u>2.3 ± 4.1</u>	<u>8.2</u>
<u>7.5 - 10.0</u>	<u>1.2 ± 0.04</u>	<u>0.002 ± 0.004</u>	<u>0.88</u>	<u>18.6 ± 6.4</u>	<u>12.8</u>
<u>10.5 - 13.0</u>	<u>1.2 ± 0.06</u>	<u>0.002 ± 0.001</u>	<u>0.92</u>	<u>27.7 ± 5.0</u>	<u>10.0</u>
<u>13.0 - 17.0</u>	<u>0.89 ± 0.08</u>	<u>0.003 ± 0.001</u>	<u>0.78</u>	<u>40.6 ± 7.2</u>	<u>14.2</u>

of NDACC (B. Langerock, personal communication, 2017), and for C_2H_6 we applied an improved version applied in Franco et al. (2015) (E. Mahieu, personal communication, 2017). Pressure and temperature profiles are from NCEP at NDACC. For the retrieval of WV we use ERA-d to imitate our typical retrieval strategy. Similar as WV, full error analysis is performed, i.e., mainly considering measurement noise error and forward model parameter errors (see Sect. 3.1).

5 The retrieval of HCN, CO, and C_2H_6 was performed only during dates with NOAA FPH sonde measurements. Since the FPH profiles are taken as the ground 'truth' we have limited spectra taken only within 1 h of the sonde launch based on findings presented earlier. In all cases, the standard settings remain the same and only the WV profile reference is changed. An example of the effect of WV profile in the retrieval of the different gases is shown in Figure 10. The different WV profiles used on this day (July 22 2014) are shown on top. The retrieved WV (black) is the closest in shape and magnitude to the NOAA FPH profile 10 (purple). All the other WV profiles show significant differences with respect to the FPH. The gas profile retrievals are shown on the left panels using similar color scheme as in the WV profile panel. The relative difference at every retrieval level, defined as $(x_i - x_{fph})/x_{fph} \times 100$, is shown on the right panels. The lowest relative difference in all grid points and for all gases is always when using the retrieved WV profile (black). All other WV sources present significant differences. For example, for HCN differences of up to -20 % are found at 6-10 km if using ERA-I. CO and C_2H_6 also show important differences but always 15 below 10 %. This example suggests that the current retrieval strategy of WV is suitable ~~for obtaining WV vertical profiles and will improve the retrieval of other to avoid WV interference in the retrievals of other trace gases.~~

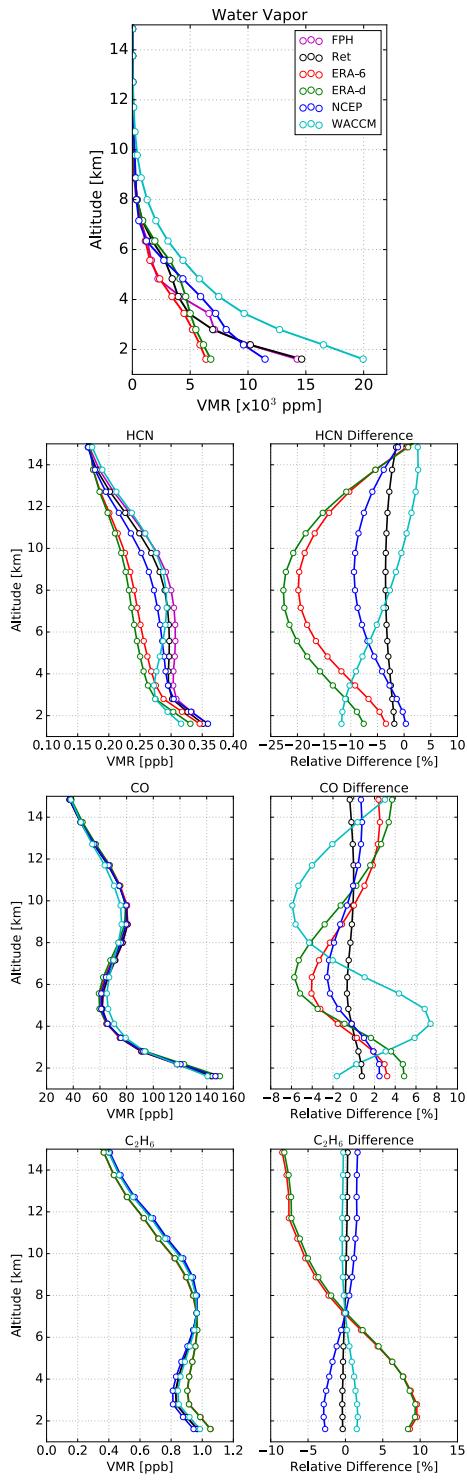


Figure 10. Example on July 22 2014 of retrieval profiles of the HCN, CO, and C_2H_6 using the different WV a priori sources shown on top. The retrieval profiles are left and right panels represent the relative different in percent with respect to the retrieval, which uses NOAA FPH WV.

Table 3. Retrieval settings of gases to study the influence of WV. All interfering species are fitted with scaling factor, except O₃ in the retrieval of CO and C₂H₆ and is fitted as vertical profile.

Gas	Micro-windows [cm ⁻¹]	Interfering species
CO	2057.7-2058.0;	O ₃ , CO ₂ , OCS, H ₂ O,
	2069.56-2069.76;	N ₂ O
	2157.50- 2159.15	
HCN	3268.04-3268.40;	H ₂ O, C ₂ H ₂ , CO ₂ , O ₃
	3287.10-3287.35;	
	3299.40-3299.60	
C ₂ H ₆	2976.66-2977.059;	
	2983.20-2983.50;	O ₃ , H ₂ O, CH ₄ , CH ₃ Cl
	2986.45-2986.85	

In order to determine the general impact of the different WV sources for all spectra recorded within 1 h of sonde launch for 6 years we have performed an ODR and statistical analysis similar to the one presented in section 4.3. In this case, the retrieval using NOAA FPH WV is used as the reference. Figure 11 shows the main results of the ODR analysis for the three gases using the different WV sources, and at different layers. The best correlations (r-value) and the lowest intercepts are found using the 5 pre-retrieved WV profiles for all three gases, in agreement with the example given in Fig. 10. The slope values are close to unity and within the uncertainty values for CO (middle) and C₂H₆ (right) using the pre-retrieved WV. However, HCN on the left shows the most notable difference with respect to unity. The intercept is normally negligible for pre-retrieved WV for all gases. The bias and precision results are shown in Fig. 12. A bias larger than 6 and 1 % are found for HCN and CO respectively using WACCM WV in the layer close to the surface. C₂H₆ does not show a significant bias among different layers and WV 10 sources. Overall, these results suggest that incorporating the pre-retrieved WV in the forward model improves the quality of other retrieved gases.

6 Conclusions

The aim of the present research was to determine the limitations to retrieve real WV structural variability from the boundary layer to the upper troposphere ~~-lower stratosphere~~ using a standard FTIR inversion, i.e., a current retrieval strategy is not 15 ~~further improved modified~~ to correlate well with reference vertical profiles. Highly precise and accurate vertical profiles of WV from NOAA balloon FPH in-situ sondes are used for the first time as reference to evaluate FTIR WV profiles in BLD and MLO allowing the characterization of the retrievals in mid-latitudes boundary layer and sub-tropical free troposphere.

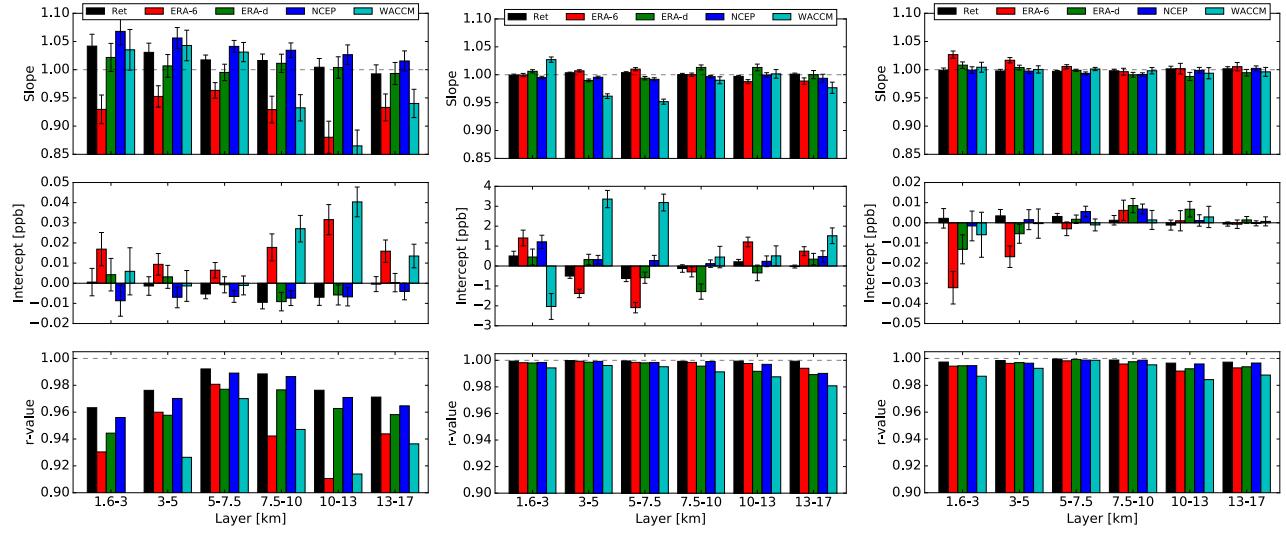


Figure 11. Results of the ODR analysis where the mixing ratios using different WV sources at different layers are compared with the 'truth' retrieved values using the NOAA FPH WV for HCN (left), CO (middle), and C_2H_6 (right). Error bars represent the standard errors of the estimated parameters.

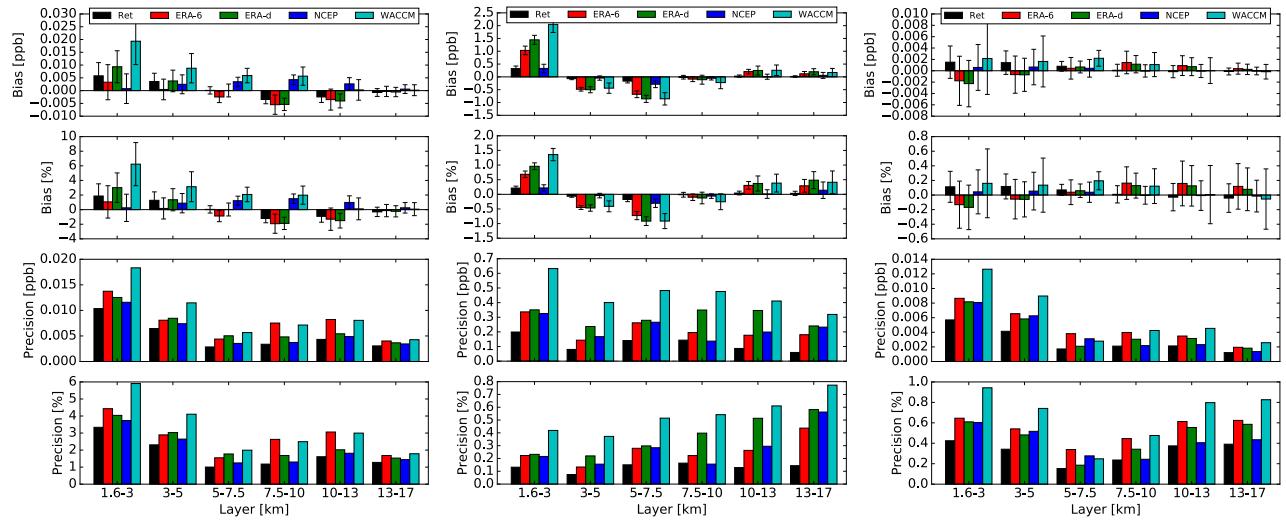


Figure 12. Statistical analysis results (bias and precision) for HCN (left), CO (middle), and C_2H_6 (right) using different WV profiles at different altitudes. The error bars in the bias represent the standard error of the distribution.

The spatial-temporal variability of WV is inferred prior to a quantitative comparison. By using daily continuous FTIR measurements we derive a temporal variability for different altitudes and find that at BLD the different layers are highly correlated and show comparable variability. In contrast, at MLO the variability among layers is quite different indicating vigorous inhomogeneity due to local convection or long-range transport. The ideal coincidence time between sonde launch 5 and FTIR measurements are 0 - 30 min and 0 - 60 min in BLD and MLO, respectively to avoid variability larger than 2 % for all altitudes. The horizontal position with maximal sensitivity of WV distribution is derived for each FTIR measurement. Then, based on the sonde location at each altitude the horizontal spatial mismatch is characterized. The insights gained from this evaluation is that the boundary layer (about 1.5 to 3 km in Boulder) is the only layer where the air mass probed by the FTIR and NOAA FPH in-situ is likely unchanged since the horizontal difference remains below 10 km. We show that 10 above 5 km the spatial mismatch increased significantly up to 60 km horizontal distance at about 10 km altitude. This feature does not depend on the coincidence time between measurements but rather on the local to synoptic meteorological scales. More broadly, even co-located FTIR and sonde launch measurements would have significant horizontal mismatch at different altitudes. Further work needs to be done to establish the best methodology to validate FTIR profile retrievals while avoiding difference in geometry of measurements.

15 This work offers a new assessment of the accuracy and precision of FTIR retrievals at different altitudes. The analysis consists of the comparison of WV ~~using for~~ several atmospheric layers using a-ODR and statistical analysis, i.e., estimation of accuracy and precision. Furthermore, we study the effect of different WV a priori commonly used among NDACC stations (ERA-I, NCEP, and WACCM ~~sources~~ ~~profiles~~) and ~~the limitations of the FTIR WV averaging kernels by comparing un-smoothed and smoothed FPH profiles with FTIR retrievals~~. The following overall conclusions can be drawn from the un- 20 smoothed comparison of WV using several layers: (1) using 6 hourly and daily ERA-I a priori shows the best correlation and comparison in both sites; (2) the lowest bias and precision are found in the closest layer to the instrument (1.5 - 3km at BLD and 3 - 5 km at MLO). At BLD, we report a negligible negative bias of $-0.001 \pm 0.105 \times 10^3$ ppm ($-0.02 \pm 1.9\%$) and precision of 0.21×10^3 ppm (3.7 %) for the 1.5 - 3 km layer while at MLO the bias is $-0.10 \pm 0.08 \times 10^3$ ppm ($-5.8 \pm 4.6\%$) and precision of 0.16×10^3 ppm (9.2 %) for the 3 - 5.5 km layer, which are larger likely due to the significant spatial mismatch difference 25 between the location of measurements; (3) high vertical variability probed by the sonde in the second layer is not fully captured by the retrievals, although is considerably better than a priori profiles; (4) and one significant findings to emerge is that the retrievals show encouraging results in the 10.5 - 13.5 km at BLD and 13 - 16 km at MLO (roughly the UTLS layer) with $13.1 \pm 5.3\%$ (BLD) bias and a precision of 10.6 % (BLD) but the bias increases to about 40 % above this layer. Table 2 was constructed to show a representative analysis when the spatial mismatch is known and when the location of the FTIR and the 30 launch of the sonde are near each other. ~~In this table results are shown for un-smoothed and smoothed FPH profiles~~. According to these results we infer that the interpretation of the averaging kernels and degrees of freedom are quite conservative and WV retrievals contain more information than expected. ~~Among all layers, the biases are lower for un-smoothed FPH profiles indicating limitations of WV averaging kernels~~. The findings of this study show that FTIR profiles can be used to evaluate long-term records of WV at several ~~unique~~ partial columns in the troposphere. ~~Further research would explore the additional WV absorption features in order to improve the information content, e.g., micro-windows employed in the latest MUSICA~~

version. Also, as we show, the ERA-I WV profiles yield lower biases, hence we would construct a priori covariance matrices for these that maximize accuracy and vertical structure.

The second goal of this study was to investigate the influence of WV in the retrieval of other tropospheric gas profiles with DOF larger than two. Here we present results for three important gases, i.e., HCN, CO, and C₂H₆ using the WV NOAA FPH profile as ground 'truth' as reference and comparing to other WV ~~sources~~profiles, including the retrieved WV, ERA-I, NCEP, and WACCM. In general, our results recommend retrieving WV profiles first then using them as input to the retrievals of other gases in order to reduce bias due to imperfect WV vertical profile. As an example (Fig. 10) we show relative differences of up to 25 % at 8 km, 8 % at 4 km, and 10 % at 3 km for HCN, CO, and C₂H₆ if WV is not retrieved beforehand and used ~~posterior~~ it as the input WV profile. Overall, a statistical comparison of all profiles in the 1.5 - 3.0 km layer show significant impact on HCN (about 6 % bias), middle impact in CO (about 1.2 % bias), and low impact on C₂H₆ (< 0.5 % bias). This sensitivity study is the first comprehensive quantitative investigation in this topic and provides a basis for future error budget assessment. In principle we hypothesize that the effect of WV profiles might be ~~bigger in humid sites~~ larger in humid regions within the boundary layer but further research should be carried out to establish its quantitative importance.

7 Data availability

15 The NCAR FTIR water vapor retrievals can be obtained from the authors upon request. Vertical Profile of Water Vapor from Balloon flight NOAA can be accessed through the website: https://www.esrl.noaa.gov/gmd/dv/data/index.php?parameter_name=Water%2BVapor.

Disclaimer. The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the 20 National Science Foundation.

Acknowledgements. This study has been supported under contract by the National Aeronautics and Space Administration (NASA). We are grateful to the NOAA staff at MLO for technical support and maintenance of the NCAR FTIR. Especially, we wish to thank Paul Fukumura. We would like to thank David Nardini and Darryl Kuniyuki for diligently preparing and launching the NOAA FPH instruments monthly from Hilo, Hawaii. We thank Helen Worden for her valuable suggestions during the NCAR internal review.

References

Barthlott, S., Schneider, M., Hase, F., Blumenstock, T., Kiel, M., Dubravica, D., García, O. E., Sepúlveda, E., Mengistu Tsidu, G., Takele Keene, S., Grutter, M., Plaza-Medina, E. F., Stremme, W., Strong, K., Weaver, D., Palm, M., Warneke, T., Notholt, J., Mahieu, E., Servais, C., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Tropospheric water vapour isotopologue data (H_2^{16}O , H_2^{18}O , and HD^{16}O) as obtained from NDACC/FTIR solar absorption spectra, *Earth System Science Data*, 9, 15–29, doi:10.5194/essd-9-15-2017, 5 <https://www.earth-syst-sci-data.net/9/15/2017/>, 2017.

Dammers, E., Shephard, M. W., Palm, M., Cady-Pereira, K., Capps, S., Lutsch, E., Strong, K., Hannigan, J. W., Ortega, I., Toon, G. C., Stremme, W., Grutter, M., Jones, N., Smale, D., Siemons, J., Hrpcek, K., Tremblay, D., Schaap, M., Notholt, J., and Erisman, J. W.: Validation of the CrIS fast physical NH_3 retrieval with ground-based FTIR, *Atmospheric Measurement Techniques*, 10, 2645–2667, 10 doi:10.5194/amt-10-2645-2017, 2017.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Källberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: 15 configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597, doi:10.1002/qj.828, 2011.

Dessler, A. E.: Cloud variations and the Earth's energy budget, *Geophysical Research Letters*, 38, doi:10.1029/2011GL049236, 119701, 2011.

Finger, F. G., Gelman, M. E., Wild, J. D., Chanin, M. L., Hauchecorne, A., and Miller, A. J.: Evaluation of NMC Upper-Stratospheric Temperature Analyses Using Rocketsonde and Lidar Data, *Bulletin of the American Meteorological Society*, 74, 789–800, doi:10.1175/1520-0477(1993)074<0789:EONUST>2.0.CO;2, [https://doi.org/10.1175/1520-0477\(1993\)074<0789:EONUST>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<0789:EONUST>2.0.CO;2), 1993.

Franco, B., Bader, W., Toon, G., Bray, C., Perrin, A., Fischer, E., Sudo, K., Boone, C., Bovy, B., Lejeune, B., Servais, C., and Mahieu, E.: Retrieval of ethane from ground-based FTIR solar spectra using improved spectroscopy: Recent burden increase above Jungfraujoch, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 160, 36 – 49, doi:<https://doi.org/10.1016/j.jqsrt.2015.03.017>, 2015.

Franco, B., Mahieu, E., Emmons, L. K., Tzompa-Sosa, Z. A., Fischer, E. V., Sudo, K., Bovy, B., Conway, S., Griffin, D., Hannigan, J. W., Strong, K., and Walker, K. A.: Evaluating ethane and methane emissions associated with the development of oil and natural gas extraction in North America, *Environmental Research Letters*, 11, 044 010, doi:10.1088/1748-9326/11/4/044010, 2016.

Galewsky, J. and Rabanus, D.: A Stochastic Model for Diagnosing Subtropical Humidity Dynamics with Stable Isotopologues of Water Vapor, *Journal of the Atmospheric Sciences*, 73, 1741–1753, doi:10.1175/JAS-D-15-0160.1, 2016.

García, O. E., Schneider, M., Hase, F., Blumenstock, T., Sepúlveda, E., and González, Y.: Quality assessment of ozone total column amounts as monitored by ground-based solar absorption spectrometry in the near infrared ($> 3000 \text{ cm}^{-1}$), *Atmospheric Measurement Techniques*, 7, 3071–3084, doi:10.5194/amt-7-3071-2014, <https://www.atmos-meas-tech.net/7/3071/2014/>, 2014.

Garcia, R. R., Marsh, D. R., Kinnison, D. E., Boville, B. A., and Sassi, F.: Simulation of secular trends in the middle atmosphere, 1950–2003, *Journal of Geophysical Research: Atmospheres*, 112, doi:10.1029/2006JD007485, d09301, 2007.

Goff, J.: Saturation pressure of water on the new Kelvin temperature scale, *Transactions of the American society of heating and ventilating engineers*, 63, 347–354, 1957.

Hall, E. G., Jordan, A. F., Hurst, D. F., Oltmans, S. J., Vömel, H., Kühnreich, B., and Ebert, V.: Advancements, measurement uncertainties, and recent comparisons of the NOAA frost point hygrometer, *Atmospheric Measurement Techniques*, 9, 4295–4310, doi:10.5194/amt-9-4295-2016, 2016.

Hannigan, J. W., Coffey, M. T., and Goldman, A.: Semiautonomous FTS observation system for remote sensing of stratospheric and tropospheric gases, *Journal of Atmospheric and Oceanic Technology*, 26, 1814–1828, doi:10.1175/2009JTECHA1230.1, 2009.

Hase, F., Hannigan, J. W., Coffey, M. T., Goldman, A., Höpfner, M., Jones, N. B., Rinsland, C. P., and Wood, S. W.: Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 87, 25–52, doi:10.1016/j.jqsrt.2003.12.008, 2004.

Held, I. M. and Soden, B. J.: Water vapor feedback and global warming, *Annual Review of Energy and the Environment*, 25, 441–475, doi:10.1146/annurev.energy.25.1.441, 2000.

Hurst, D. F., Hall, E. G., Jordan, A. F., Miloshevich, L. M., Whiteman, D. N., Leblanc, T., Walsh, D., Vömel, H., and Oltmans, S. J.: Comparisons of temperature, pressure and humidity measurements by balloon-borne radiosondes and frost point hygrometers during MOHAVE-2009, *Atmospheric Measurement Techniques*, 4, 2777–2793, doi:10.5194/amt-4-2777-2011, 2011a.

Hurst, D. F., Oltmans, S. J., Vömel, H., Rosenlof, K. H., Davis, S. M., Ray, E. A., Hall, E. G., and Jordan, A. F.: Stratospheric water vapor trends over Boulder, Colorado: Analysis of the 30 year Boulder record, *Journal of Geophysical Research Atmospheres*, 116, 1–12, doi:10.1029/2010JD015065, 2011b.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, *Bulletin of the American Meteorological Society*, 77, 437–472, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Kiehl, J. T. and Trenberth, K. E.: Earth's Annual Global Mean Energy Budget, *Bulletin of the American Meteorological Society*, 78, 197–208, doi:10.1175/1520-0477(1997)078<0197:EAGMEB>2.0.CO;2, https://doi.org/10.1175/1520-0477(1997)078<0197:EAGMEB>2.0.CO;2, 1997.

Kille, N., Baidar, S., Handley, P., Ortega, I., Sinreich, R., Cooper, O. R., Hase, F., Hannigan, J. W., Pfister, G., and Volkamer, R.: The CU mobile Solar Occultation Flux instrument: structure functions and emission rates of NH₃, NO₂ and C₂H₆, *Atmospheric Measurement Techniques*, 10, 373–392, doi:10.5194/amt-10-373-2017, 2017.

Kinnison, D. E., Brasseur, G. P., Walters, S., Garcia, R. R., Marsh, D. R., Sassi, F., Harvey, V. L., Randall, C. E., Emmons, L., Lamarque, J. F., Hess, P., Orlando, J. J., Tie, X. X., Randel, W., Pan, L. L., Gettelman, A., Granier, C., Diehl, T., Niemeier, U., and Simmons, A. J.: Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model, *Journal of Geophysical Research: Atmospheres*, 112, doi:10.1029/2006JD007879, d20302, 2007.

Korn, G. and Korn, T.: *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*, Dover Civil and Mechanical Engineering Series, Dover Publications, https://books.google.com/books?id=xHNd5zCXt-EC, 2000.

Kulawik, S. S., Bowman, K. W., Luo, M., Rodgers, C. D., and Jourdain, L.: Impact of nonlinearity on changing the a priori of trace gas profile estimates from the Tropospheric Emission Spectrometer (TES), *Atmospheric Chemistry and Physics*, 8, 3081–3092, doi:10.5194/acp-8-3081-2008, 2008.

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., and Polvani, L. M.: Climate Change from 1850 to 2005 Simulated in CESM1(WACCM), *Journal of Climate*, 26, 7372–7391, doi:10.1175/JCLI-D-12-00558.1, 2013.

Noone, D.: Pairing Measurements of the Water Vapor Isotope Ratio with Humidity to Deduce Atmospheric Moistening and Dehydration in the Tropical Midtroposphere, *Journal of Climate*, 25, 4476–4494, doi:10.1175/JCLI-D-11-00582.1, <https://doi.org/10.1175/JCLI-D-11-00582.1>, 2012.

Oltmans, S. J., Vömel, H., Hofmann, D. J., Rosenlof, K. H., and Kley, D.: The increase in stratospheric water vapor from balloon-borne, frostpoint hygrometer measurements at Washington, D.C., and Boulder, Colorado, *Geophysical Research Letters*, 27, 3453–3456, doi:10.1029/2000GL012133, 2000.

Paton-Walsh, C., Deutscher, N. M., Griffith, D. W. T., Forgan, B. W., Wilson, S. R., Jones, N. B., and Edwards, D. P.: Trace gas emissions from savanna fires in northern Australia, *Journal of Geophysical Research: Atmospheres*, 115, doi:10.1029/2009JD013309, d16314, 2010.

Pougatchev, N. S., Connor, B. J., and Rinsland, C. P.: Infrared measurements of the ozone vertical distribution above Kitt Peak, *Journal of Geophysical Research: Atmospheres*, 100, 16 689–16 697, doi:10.1029/95JD01296, 1995.

Rinsland, C. P., Jones, N. B., Connor, B. J., Logan, J. A., Pougatchev, N. S., Goldman, A., Murcray, F. J., Stephen, T. M., Pine, A. S., Zander, R., Mahieu, E., and Demoulin, P.: Northern and southern hemisphere ground-based infrared spectroscopic measurements of tropospheric carbon monoxide and ethane, *Journal of Geophysical Research: Atmospheres*, 103, 28 197–28 217, doi:10.1029/98JD02515, 1998.

Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific, Singapore, 2000.

Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, *Journal of Geophysical Research (Atmospheres)*, 108, 4116, doi:10.1029/2002JD002299, 2003.

Rothman, L., Gordon, I., Barbe, A., Benner, D., Bernath, P., Birk, M., Boudon, V., Brown, L., Campargue, A., Champion, J.-P., Chance, K., Coudert, L., Dana, V., Devi, V., Flaud, J.-M., Gamache, R., Goldman, A., Jacquemart, D., Kleiner, I., Lacome, N., Lafferty, W., Mandin, J.-Y., Massie, S., Mikhailenko, S., Miller, C., Moazzen-Ahmadi, N., Naumenko, O., Nikitin, A., Orphal, J., Perevalov, V., Perrin, A., Predoi-Cross, A., Rinsland, C., Rotger, M., Šimečková, M., Smith, M., Sung, K., Tashkun, S., Tennyson, J., Toth, R., Vandaele, A., and Auwera, J. V.: The HITRAN 2008 molecular spectroscopic database, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 110, 533 – 572, doi:<https://doi.org/10.1016/j.jqsrt.2009.02.013>, hITRAN, 2009.

Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Bizzocchi, L., Boudon, V., Brown, L. R., Campargue, A., Chance, K., Cohen, E. A., Coudert, L. H., Devi, V. M., Drouin, B. J., Fayt, A., Flaud, J. ., Gamache, R. R., Harrison, J. J., Hartmann, J. ., Hill, C., Hodges, J. T., Jacquemart, D., Jolly, A., Lamouroux, J., Le Roy, R. J., Li, G., Long, D. A., Lyulin, O. M., Mackie, C. J., Massie, S. T., Mikhailenko, S., Mueller, H. S. P., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E. R., Richard, C., Smith, M. A. H., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G. C., Tyuterev, V. G., and Wagner, G.: The HITRAN 2012 molecular spectroscopic database, *Journal of Quantitative Spectroscopy & Radiative Transfer*, 130, 4–50, doi:10.1016/j.jqsrt.2013.07.002, 2013.

Scherer, M., Vömel, H., Fueglister, S., Oltmans, S. J., and Staehelin, J.: Trends and variability of midlatitude stratospheric water vapour deduced from the re-evaluated Boulder balloon series and HALOE, *Atmospheric Chemistry and Physics*, 8, 1391–1402, doi:10.5194/acp-8-1391-2008, 2008.

Schneider, M. and Hase, F.: Ground-based FTIR water vapour profile analyses, *Atmospheric Measurement Techniques*, 2, 609–619, doi:10.5194/amt-2-609-2009, <https://www.atmos-meas-tech.net/2/609/2009/>, 2009.

Schneider, M., Hase, F., and Blumenstock, T.: Water vapour profiles by ground-based FTIR spectroscopy: study for an optimised retrieval and its validation, *Atmospheric Chemistry and Physics*, 6, 811–830, doi:10.5194/acp-6-811-2006, 2006.

Schneider, M., Yoshimura, K., Hase, F., and Blumenstock, T.: The ground-based FTIR network's potential for investigating the atmospheric water cycle, *Atmospheric Chemistry and Physics*, 10, 3427–3442, doi:10.5194/acp-10-3427-2010, <https://www.atmos-chem-phys.net/10/3427/2010/>, 2010.

Schneider, M., Barthlott, S., Hase, F., González, Y., Yoshimura, K., García, O. E., Sepúlveda, E., Gomez-Pelaez, A., Gisi, M., Kohlhepp, R., Dohe, S., Blumenstock, T., Wiegele, A., Christner, E., Strong, K., Weaver, D., Palm, M., Deutscher, N. M., Warneke, T., Notholt, J., Lejeune, B., Demoulin, P., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Ground-based remote sensing of tropospheric water vapour isotopologues within the project MUSICA, *Atmospheric Measurement Techniques*, 5, 3007–3027, doi:10.5194/amt-5-3007-2012, 2012.

Schneider, M., Wiegele, A., Barthlott, S., González, Y., Christner, E., Dyroff, C., García, O. E., Hase, F., Blumenstock, T., Sepúlveda, E., Mengistu Tsidu, G., Takele Kenea, S., Rodríguez, S., and Andrey, J.: Accomplishments of the MUSICA project to provide accurate, long-term, global and high-resolution observations of tropospheric $\{\text{H}_2\text{O}, \delta\text{D}\}$ pairs – a review, *Atmospheric Measurement Techniques*, 9, 2845–2875, doi:10.5194/amt-9-2845-2016, <https://www.atmos-meas-tech.net/9/2845/2016/>, 2016.

Seinfeld, J. H. and Pandis, S. N.: *Atmospheric Chemistry and Physics: from air pollution to climate change*, Wiley-Interscience, Hoboken, New Jersey, 2 edn., 2006.

Sepúlveda, E., Schneider, M., Hase, F., Barthlott, S., Dubravica, D., García, O. E., Gomez-Pelaez, A., González, Y., Guerra, J. C., Gisi, M., Kohlhepp, R., Dohe, S., Blumenstock, T., Strong, K., Weaver, D., Palm, M., Sadeghi, A., Deutscher, N. M., Warneke, T., Notholt, J., Jones, N., Griffith, D. W. T., Smale, D., Brailsford, G. W., Robinson, J., Meinhardt, F., Steinbacher, M., Aalto, T., and Worthy, D.: Tropospheric CH_4 signals as observed by NDACC FTIR at globally distributed sites and comparison to GAW surface in situ measurements, *Atmospheric Measurement Techniques*, 7, 2337–2360, doi:10.5194/amt-7-2337-2014, <https://www.atmos-meas-tech.net/7/2337/2014/>, 2014.

Suortti, T. M., Kats, A., Rivi, R., Kämpfer, N., Leiterer, U., Miloshevich, L. M., Neuber, R., Paukkunen, A., Ruppert, P., Vömel, H., and Yushkov, V.: Tropospheric comparisons of Vaisala radiosondes and balloon-borne frost-Point and Lyman- α hygrometers during the LAUTLOS-WAVVAP experiment, *Journal of Atmospheric and Oceanic Technology*, 25, 149–166, doi:10.1175/2007JTECHA887.1, 2008.

Sussmann, R. and Borsdorff, T.: Technical Note: Interference errors in infrared remote sounding of the atmosphere, *Atmospheric Chemistry and Physics*, 7, 3537–3557, doi:10.5194/acp-7-3537-2007, <https://www.atmos-chem-phys.net/7/3537/2007/>, 2007.

Sussmann, R., Borsdorff, T., Rettinger, M., Camy-Peyret, C., Demoulin, P., Duchatelet, P., Mahieu, E., and Servais, C.: Technical Note: Harmonized retrieval of column-integrated atmospheric water vapor from the FTIR network-first examples for long-term records and station trends, *Atmospheric Chemistry and Physics*, 9, 8987–8999, doi:10.5194/acp-9-8987-2009, 2009.

Trenberth, K. E. and Asrar, G. R.: Challenges and Opportunities in Water Cycle Research: WCRP Contributions, *Surveys in Geophysics*, 35, 515–532, doi:10.1007/s10712-012-9214-y, 2014.

Tzompa-Sosa, Z. A., Mahieu, E., Franco, B., Keller, C. A., Turner, A. J., Helmig, D., Fried, A., Richter, D., Weibring, P., Walega, J., Yacovitch, T. I., Herndon, S. C., Blake, D. R., Hase, F., Hannigan, J. W., Conway, S., Strong, K., Schneider, M., and Fischer, E. V.: Revisiting global fossil fuel and biofuel emissions of ethane, *Journal of Geophysical Research: Atmospheres*, 122, 2493–2512, doi:10.1002/2016JD025767, 2016.

Viatte, C., Strong, K., Walker, K. A., and Drummond, J. R.: Five years of CO , HCN , C_2H_6 , C_2H_2 , CH_3OH , HCOOH and H_2CO total columns measured in the Canadian high Arctic, *Atmospheric Measurement Techniques*, 7, 1547–1570, doi:10.5194/amt-7-1547-2014, 2014.

Vigouroux, C., Stavrakou, T., Whaley, C., Dils, B., Duflot, V., Hermans, C., Kumps, N., Metzger, J.-M., Scolas, F., Vanhaelewijn, G., Müller, J.-F., Jones, D. B. A., Li, Q., and De Mazière, M.: FTIR time-series of biomass burning products (HCN, C₂H₆, C₂H₂, CH₃OH, and HCOOH) at Reunion Island (21 S, 55 E) and comparisons with model data, *Atmospheric Chemistry and Physics*, 12, 10 367–10 385, doi:10.5194/acp-12-10367-2012, 2012.

5 Vigouroux, C., Blumenstock, T., Coffey, M., Errera, Q., García, O., Jones, N. B., Hannigan, J. W., Hase, F., Liley, B., Mahieu, E., Mellqvist, J., Notholt, J., Palm, M., Persson, G., Schneider, M., Servais, C., Smale, D., Thölix, L., and De Mazière, M.: Trends of ozone total columns and vertical distribution from FTIR observations at eight NDACC stations around the globe, *Atmospheric Chemistry and Physics*, 15, 2915–2933, doi:10.5194/acp-15-2915-2015, 2015.

Vogelmann, H., Sussmann, R., Trickl, T., and Reichert, A.: Spatiotemporal variability of water vapor investigated using lidar and FTIR vertical soundings above the Zugspitze, *Atmospheric Chemistry and Physics*, doi:10.5194/acp-15-3135-2015, 2015.

10 Wild, J. D., Gelman, M. E., Miller, A. J., Chanin, M. L., Hauchecorne, A., Keckhut, P., Farley, R., Dao, P. D., Meriwether, J. W., Gobbi, G. P., Congeduti, F., Adriani, A., McDermid, I. S., McGee, T. J., and Fishbein, E. F.: Comparison of stratospheric temperatures from several lidars, using National Meteorological Center and microwave limb sounder data as transfer references, *Journal of Geophysical Research: Atmospheres*, 100, 11 105–11 111, doi:10.1029/95JD00631, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95JD00631>, 1995.

15 Wu, C. and Yu, J. Z.: Evaluation of linear regression techniques for atmospheric applications: the importance of appropriate weighting, *Atmospheric Measurement Techniques*, 11, 1233–1250, doi:10.5194/amt-11-1233-2018, 2018.