TABLE OF CONTENTS

Comments from Reviewer #1

Box 1.1

Review #1
Anonymous during peer-review: Yes
Anonymous in acknowledgements of published article: Yes

Recommendation to the editor
1) Scientific significance
Does the manuscript represent a substantial contribution to scientific progress within the scope of this journal (substantial new concepts, ideas, methods, or data)? Good

2) Scientific quality
Are the scientific approaches and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)? Note that papers do not necessarily need to be long to be scientifically sound. Fair

3) Presentation quality
Are the scientific results and conclusions presented in a clear, concise, and well structured way (number and quality of figures/tables, appropriate use of English language)? Good

For final publication, the manuscript should be
accepted subject to minor revisions

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

We thank the reviewer for providing encouraging comments.

Box 1.2

I agree with the comments of the other reviewer that this seems more like a modeling paper than a measurements paper with only a little description of the instrument characteristics and a very small number of measurements. Error estimates are missing.

We thank the reviewer for this very constructive comment. We would like to point out that this is the first day/night contiguous measurement of NO_2 column being published, which provides an important test of photochemical models. Our pilot study with a week of data successfully demonstrates the feasibility of retrieving both daytime and nighttime NO_2 abundances using the grating spectrometer measurements and the applicability of the modified Langley method, as well as the qualitative agreement with the photochemical model. In the last review cycle, we added a new Figure 2 showing the spectral fit and an inset of the new Figure 3 that shows the errors of the QDOAS fitting. We also added the signal-to-noise ratios of the measurements and added the uncertainty of the parameters of the Langley extrapolation. The new information has been presented in the revised manuscript but was not present in the original discussion paper published on the AMTD website. We hope that the new information provided in the last review cycle has addressed this comment. We are happy to provide more information if the reviewer would like to see additional error estimates.

Box 1.3

Even with intrusion from nearby cities, the NO_2 amounts are very low on the "worst" day, October 27, about 0.2 DU, barely above stratospheric values. This paper is totally about the stratospheric behavior of NO_2 and should be described as such.

We greatly appreciate this comment. We have replaced "total column NO_2" by "stratospheric column NO_2" in most places (including the title), except when we discussed the measurement on October 27, 2018 when the tropospheric contribution was significant, we simply call it "NO_2". We hope that these changes would help clarify our measurements.

Box 1.4

Even with intrusion from nearby cities, the NO2 amounts are very low on the "worst" day, October 27, about 0.2 DU, barely above stratospheric values. This paper is totally about the stratospheric behavior of NO2 and should be described as such. The 1D modeling is adequate for stratospheric behavior and is useful as an indicator that the measurements are reasonable. The non-linear behavior compared to the model during the daytime is not explained either as a chemistry result or as an instrument problem. Aside from October 27, the curvature is repeatable on multiple measurements (Figs. 3 and 6) and should be explained if it is possibly an instrument problem.

We believe that the reviewer is referring to Figures 3 and 6 of the original discussion paper published on the AMTD website. In the revised manuscript, the corresponding figures are Figures 4 and 7.

We greatly appreciate this comment. Since the Editor has the same concern, we reinvestigated our Langley extrapolation. We found that the curvature of the daytime data was likely due to two artifacts (please see Box E.2 on page 7 of this response letter for a more detailed discussion): (1) a low percentile that defines the baseline of the Langley plot and (2) the bias of the extrapolation due to the sparse data points at high air mass factors. After correcting these artifacts, the curvature is significantly reduced. We thus believe that the curvature is unlikely due to an instrument problem. We have modified our Figures 3, 4, 7, A1 and B1 accordingly.

Box 1.5

The new figures and some of the discussion given in the reply to reviewers must be part of this paper before publication.

We believe that the reviewer is referring to the new figures we showed in our previous response letter. Those figures were included in the last revised manuscript. The original discussion paper published on the AMTD website does not have those new figures. We are happy to provide direct access to the revised manuscripts if necessary.

Box 2.1

Review #2

Anonymous during peer-review: Yes

Anonymous in acknowledgements of published article: Yes

Recommendation to the editor

1) Scientific significance

Does the manuscript represent a substantial contribution to scientific progress within the scope of this journal (substantial new concepts, ideas, methods, or data)? Good

2) Scientific quality

Are the scientific approaches and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)? Note that papers do not necessarily need to be long to be scientifically sound. Good

3) Presentation quality

Are the scientific results and conclusions presented in a clear, concise, and well structured way (number and quality of figures/tables, appropriate use of English language)? Good

For final publication, the manuscript should be
accepted subject to technical corrections

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

We greatly appreciate the reviewer for the positive evaluation of our manuscript.

Box 2.2

There are a few minor things that need some clarification:

- 3rd order polynomial used for offset correction in DOAS fitting is higher than typical (1st order). The offset is mainly used to correct spectra for instrumental stray light and 3rd order seams rather high for this wavelength range.

We thank for this thoughtful comment. We actually tested our algorithm with a linear baseline. We found that the 3rd order polynomial gave a smaller residual than the linear baseline while it does not overfit the narrow NO_2 absorption features in our spectral window. Regarding the order of the polynomial, some studies, such as Herman et al. (2009) who also retrieved NO_2 column from DOAS measurements, use 4th or higher order polynomials for wider spectral windows. Our choice of the 3rd order polynomial is a good compromise between residual reduction and overfitting.

In response to this comment, we have added the following statement in Line 121:

> "Some studies, such as Herman et al. (2009), use 4th or higher order polynomials for wider spectral windows. Since the NO_2 absorption features are much narrower than our spectral window (430–468 nm), the broad shape of the 3rd order polynomial does not affect the NO_2 retrievals. In addition, for our spectral window, we tested our retrieval algorithm using a linear baseline and we concluded that a 3rd order polynomial reduces the residuals more effectively than a linear baseline."

Box 2.3

description how NO2 mol. absorption cross section at an effective temperature is created from the climatological temperature profiles and standard NO2 profile should be improved. It will be good to state what that final effective NO2 temperature is for that location.

Our NO_2 cross section reference assumes the yearly average temperature profile at TMF and a low level of free tropospheric NO_2, thus our effective temperature is 231 K. To test the sensitivity of these assumptions we considered two extreme cases. A cooler atmosphere with a lower partition of NO_2 in the free trop and a warmer atmosphere with a higher partition of NO_2 in the free troposphere. The effective temperatures of these two cases are estimated by 229 K and 249 K, respectively. The difference between retrievals using these extreme cases is ~5%; the regular variation of temperature and tropospheric NO_2 at TMF is well within estimates.

In response to this comment, we have added the following statement in Line 126:

> "Our NO_2 cross section reference assumes the yearly average temperature profile at TMF and a low level of free tropospheric NO_2. The effective temperature of the NO_2 absorption cross section used in the work is 231 K. To test the sensitivity of these assumptions we considered two extreme cases: (i) a cooler atmosphere with a lower partition of NO_2 in the free troposphere and (ii) a warmer atmosphere with a higher partition of NO_2 in the free troposphere. The effective temperatures of these two cases are estimated by 229 K and 249 K, respectively. The difference between retrievals using these extreme cases is ~5%; the regular variation of temperature and tropospheric NO_2 at TMF is well within estimates."

Box 2.4

QDOAS version should be stated.

For this work we have used the QDOAS 3.2 (2017). We have modified the following statement in Line 115 from

> "The spectral fitting is accomplished through the Marquardt-Levenberg minimization using QDOAS retrieval software (http://uv-vis.aeronomie.be/software/QDOAS/)"

to

"The spectral fitting is accomplished through the Marquardt-Levenberg minimization using QDOAS 3.2 (released in September 2017) retrieval software (http://uv-vis.aeronomie.be/software/QDOAS/)"

Box 2.4

Figure 2 shows an example of the DOAS fit, but it is does not indicate at what time and date the spectrum and the reference were taken. Is this solar or lunar data fit? It might be interesting to see an example of both, since the illumination of the instrument changes and might impact the quality of the fit.

The sample spectral fit shown in Figure 2 was a lunar measurement at 7:25 PM on October 24, 2018, corresponding to an air mass factor of 2.21. This information can now be found in the caption of Figure 2 of the latest revised manuscript. Solar measurements generally have better signal-to-noise ratios, so we believe that the lunar sample in Figure 2 is very representative of the quality of our spectral fit. For clarity of the paper, we intend to include only 1 figure of spectral fit. If there is a strong interest in a similar spectral fit of a solar measurement, we are more than happy to add another figure in the final version of the manuscript.

Box 2.5

I recommend rephrasing: Other techniques, such as balloon-based in situ measurements (May and Webster, 1990; Moreau et al., 2005), balloon-based solar occultations (Camy-Peyret, 1995), as well as ground-based Differential Optical Absorption Spectroscopy: MAX-DOAS (Hönninger et al., 2004; Sanders et al., 1993), Direct Sun DOAS (Herman et al., 2009; Spinei et al., 2014, more updated references here) that have been actively applied in NDACC and the Pandonia Global Network. The DOAS techniques have also been employed to further characterize the vertical distributions of NO2 (Kreher et al., 2020).

We greatly appreciate this suggestion. We have modified the statement in Line 50 accordingly.

Box E.1

Editor's comment

Dear authors,

there are 2 anonymous reveiws which you should carefully address in all points. In particular, I share the view that it is not a total column paper but your results are only relevant for the stratosphere and this should be addressed via a changed title.

We thank Reviewer #1 and this Editor for this suggestion. We have modified the title accordingly.

Box E.2

Furthermore, please carefully address these additional major concerns from my side:

- I do not agree with your arguments in reply to Box 1.5 The linear fit in Figure 6 does not mean much, other than as a baseline, as there are two linear regimes, one from 07:00 to 13:00 and from 13:00 to 16:00 hours. Is there an explanation for the two regimes?
- Fig. 3a in Sussmann et al. 2005 does not show indications for any measurable non-linear diurnal change
- Your modeled diurnal variation is much closer to a linear behavior than your measurement results. This means, obviously the main reason for your observed non-linear behavior is a different (probably experimental) one, i.e., your trying to explain via modeling is flawed or can at best be used as a minor, partial explanation. So you should thoroughly discuss possible reasons which could make the measurements erroneously look non-linear, i.e., please quantitatively estimate possible measurement artifacts like airmass/zenith angle dependencies.

We would like to thank the Editor for providing this critical comment. We agree with the Editor's observations about the difference between our data and his own data. We thank the Editor for his patience during the COVID-19 pandemics and his willingness to handle our revisions and improve our manuscript.

We spent a long time to investigate the potential factors in our spectral retrievals and the Langley extrapolation that may have caused the artifacts in the diurnal cycle. We found that two pre-processings of the Langley extrapolation were likely the causes.

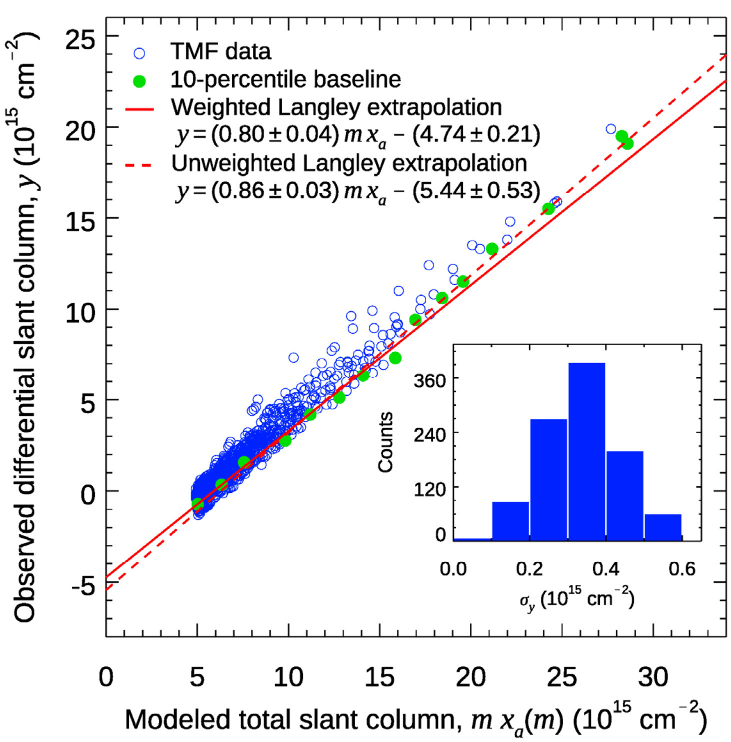
(i) A low percentile for determining the baseline

In our previous manuscript, we used a 5-percentile to define the baseline of the data cluster in Figure 3, which was then used as a clean-atmosphere baseline for the modified Langley method. However, this 5-percentile is too low compared with the 10% uncertainty of the QDOAS spectral retrieval. Thus the 5-percentile appears to be inconsistent with the uncertainty of the data points in Figure 3 and leads to uncertainty in the Langley extrapolation.

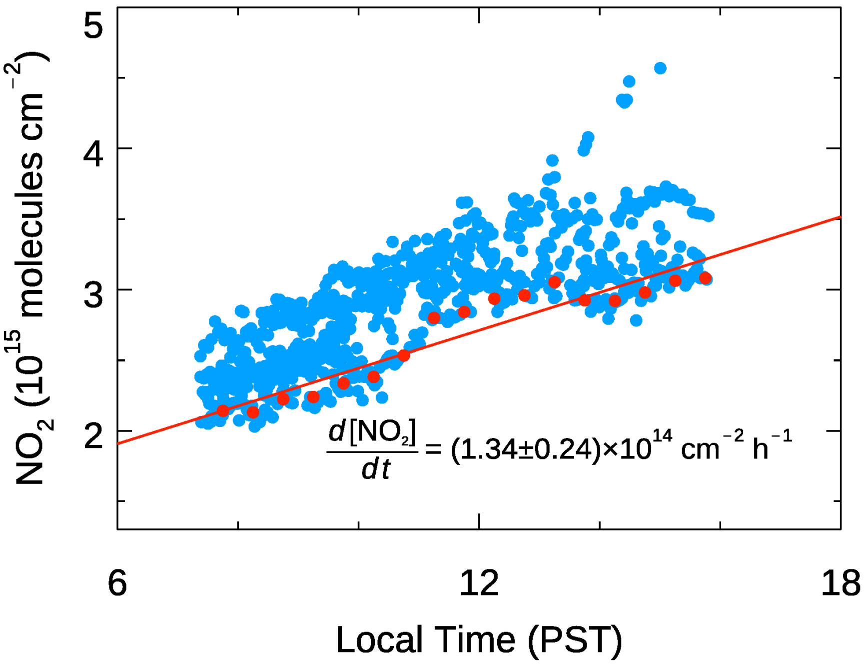
For the above reason, we have instead used a 10-percentile to define the baseline in the latest revision.

(ii) The bias in the Langley extrapolation due to sparse data points at high air mass factors

The accuracy of the Langley extrapolation critically depends on the accuracy of the diurnal baseline obtained above. Statistically, at least more than 10 data points in a bin are required to determine the 10-percentile. While we made temporally dense measurements during daytime and nighttime (at intervals less than ~20 minutes, as discussed in §2.1), the application of the air mass factor $m = \sec\theta$ in both the *x*- and *y*-axes of Figure 3 significantly stretch the time intervals at high air mass factors. The number of data points in the bins thus drops progressively by a factor of ~2: the data counts drop exponentially from 431 in the first bin, (4.5–6)×10^{15} molecules cm^{-2}, to only 12 in the 8th bin, (1.5–1.65)×10^{16} molecules cm^{-2}. The determination of the 10-percentile for bins with centers greater than 1.5×10^{16} molecules cm^{-2} is then subject to large uncertainties. Since mathematically, the 10-percentiles at high air mass factors (i.e. at the edge of the data distribution) have higher effects on a linear fit, the resultant Langley extrapolation would be strongly biased by the uncertainties of the 10-percentiles at high mass factors. Thus, to obtain a linear fit for the Langley extrapolation, we apply more weights to bins with more data counts. This definition of the weights should mimic the reduction of the variance of a sample mean by the factor of $\frac{1}{N}$ (or $\frac{1}{\sqrt{N}}$ for the standard deviation of a sample mean). Therefore, we define the weight as unity for the first bin, (4.5–6)×10^{15} molecules cm^{-2}. The weight for the second bin is the ratio of the data counts of this bin over the first bin. The weight for the third bin is the ratio of the data counts of this bin over the second bin, etc. The weighted linear fit obtained using these weights is used for the Langley extrapolation. The new Figure 3 (shown below) compares the Langley extrapolations using the weighted (solid red line) and unweighted linear fit (dashed red line). Since the 10-percentiles at high air mass (≥ 1.5×10^{16} molecules cm^{-2}) are generally overestimated due to insufficient data counts, the unweighted linear fit tends to have a steeper slope, leading to a ~15% higher reference column (5.44×10^{15} molecules cm^{-2}) relative to that obtained using the weighted linear fit. This overestimation of the reference column created the artifact in the diurnal cycle due to the normalization factor $m^{-1}(y + y_0)$.



After correcting the above pre-processing, we obtained a reference column of 4.74×10^{15} molecules cm^{-2}. We revised Figures 3, 4, 7, A1 and B1 using the corrected reference column. We specifically reproduce Figure 7 below, which estimates the linear diurnal increase rate:



The curvature is significantly reduced, and the diurnal cycle is much more linear. Our result shows that the shape of the NO_2 diurnal profile is very sensitive to the extrapolated intercept of the Langley plot, which has not been emphasized in the literature. We hope that the corrected results have addressed the concerns by the Editor and Reviewer 1.

In response to this comment, we have replaced the 5-percentiles used in previous Figures 3 and A1 by the 10-percentiles in the latest revision. We have added the following paragraph in Line 195 of the latest revision:

> "Note that the data points are sparsely distributed at high air mass factors in Figure 3. This is because while the measurements were made at relatively uniform time intervals, the air mass factor $m = sec\theta$ effectively stretch the time intervals at high air mass factors. The number of data points in the bins drops progressively by a factor of ~2: the data counts drop exponentially from 431 in the first bin, $(4.5–6)\times10^{15}$ molecules cm^{-2}, to only 12 in the bin $(1.5–1.65)\times10^{16}$ molecules cm^{-2}. The determination of the 10-percentile for bins with centers greater than 1.5×10^{16} molecules cm^{-2} is then subject to large uncertainties. Since mathematically, the 10-percentiles at high air mass factors (i.e. at the edge of the data distribution) have higher effects on a linear fit, the resultant Langley extrapolation would be strongly biased by the uncertainties of the 10-percentiles at high mass factors. Thus, to obtain a linear fit for the Langley extrapolation, we apply more weights to bins with more data counts. This definition of the weights should mimic the reduction of the variance of a sample mean by the factor of $\frac{1}{N}$ (or $\frac{1}{\sqrt{N}}$ for the standard deviation of a sample mean).

Therefore, we define the weight as unity for the first bin, $(4.5–6)\times10^{15}$ molecules cm^{-2}. The weight for the second bin, $(6–7.5)\times10^{15}$ molecules cm^{-2} is the ratio of the data counts of this bin over the first bin. The weight for the third bin is the ratio of the data counts of this bin over the second bin, etc. The weighted linear fit obtained using these weights is used for the Langley extrapolation. Figure 3 compares the Langley extrapolations using the weighted (solid red line) and unweighted linear fit (dashed red line). Since the 10-percentiles at high air mass ($\geq 1.5\times10^{16}$ molecules cm^{-2}) are generally overestimated due to insufficient data counts, the unweighted linear fit tends to have a steeper slope, leading to a ~15% higher reference column (5.44×10^{15} molecules cm^{-2}) relative to the weighted linear fit. This overestimation of the reference column may create an artifact in the diurnal cycle due to the normalization factor $m^{-1}(y + y_0)$."