Response to Referee #2:

We thank referee #2 for their helpful comments. Our responses are given below in black with the referee's comments in blue. The new text in the modified manuscript is given in red (italicized).

Referee #2:

General comments

This study presents a new approach to extend direct-sun NO2 measurements from Pandora instruments with a zenith-sky mode also applicable under cloudy conditions. In addition, attempts are made to also derive surface concentration estimates from total column observations. The general methodology is strongly inspired from the empirical zenith-sky mode developed for total ozone measurements by Dobson and Brewer spectrophotometers. Although the adopted approach implies many approximations not always well described or even identified (see detailed comments below), results are surprisingly good and certainly of interest for the ACP readership. I found the manuscript well written, concise and easy to read; also figures are of good quality and adequate in number and the appendices provide useful additional information. With one exception, credit to existing literature is appropriate. I therefore recommend publication in ACP, after careful consideration for the comments and suggestions below.

Specific comments

Section 2.1.2, L. 20: what was the temperature used for the NO2 cross-section in the zenith-sky QDOAS retrievals? Is it consistent with the effective temperature assumed for the direct-sun retrieval (254.5 K)?

The effective temperatures for NO_2 and ozone have now been included in the manuscript. The temperature used for the ZS NO_2 was 254.5 K, which is consistent with the effective temperature used for the direct-sun retrieval.

Cross sections of NO_2 at an effective temperature of 254.5 K (Vandaele et al., 1998), ozone at an effective temperature of 223 K (Bogumil et al., 2003), H_2O (Rothman et al., 2005), O_4 (Hermans et al., 2003), and Ring (Chance and Spurr, 1997) are all fitted; a fifth-order polynomial and a first-order linear offset are also included in the DOAS analysis.

Section 2.1.2, L. 25: how was the NO2 residual amount in the reference spectrum determined here? Generally speaking, this paper lacks a proper analysis of the uncertainties. It would be useful to add a section describing the estimated uncertainties for the zenith-sky column and surface concentrations (which are new data products introduced in this study).

The residual amount in the reference spectrum (RCD) was determined by using the proposed multi-non-linear regressions (see Appendix A). In short, the RCD was retrieved along with the empirical AMFs.

From the multi-non-linear model, the standard fitting error of RCD is 0.01 DU. This information was included in Section 3.1.

The RCD value used in the retrievals is 0.39 ± 0.01 DU, which is retrieved along with $AMF_{ZS\text{-}Emp}$ (Appendix A).

In addition, as suggested by the referee the following new section (Appendix D) has been included that describes the estimated uncertainties for the zenith-sky column and surface concentrations:

D. Uncertainty estimation

The uncertainties of retrieved Pandora zenith-sky NO_2 data products (total column and surface concentration) are estimated and discussed here to assess the quality of the data products. The uncertainties of total column and surface concentrations are estimated first using the uncertainty propagation method (referred to here as the UP method) based on Eqns. 2 and 3. The combined uncertainties of total column can be calculated using:

$$\sigma_{VCD_{ZS}} = \sqrt[2]{\left(\frac{\sigma_{dSCD}}{AMF}\right)^2 + \left(\frac{\sigma_{RCD}}{AMF}\right)^2 + \left(\frac{\sigma_{AMF} \times SCD}{AMF^2}\right)^2}, \qquad (7)$$

where σ_{dSCD} is the statistical uncertainty on the DOAS fit (output of QDOAS) and σ_{RCD} and σ_{AMF} are the estimated statistical uncertainties using standard errors of the RCD and the zenith-sky empirical AMF regression, respectively (Eqn. 4). To estimate the upper limit of the nominal uncertainty, AMF and SCD are used as median and maximum values in the dataset, respectively.

The combined uncertainties of the surface concentration can be calculated using:

$$\sigma_{C_{Pan}} = \sqrt[2]{\left(R_{CV}\sigma_{V_{Pan}}\right)^2 + \left(R_{CV}\sigma_{V_{strat}}\right)^2 + \left(R_{CV}\sigma_{V_{ftrop}}\right)^2 + \left(V_{Pan} - V_{strat} - V_{ftrop}\right)^2 \sigma_R^2}, \qquad (8)$$

where σ_{Vpan} is the uncertainty of Pandora zenith-sky total column NO_2, (here we use the derived σ_{VCD} in Eqn. 7), σ_{Vstrat} is the uncertainty of the stratospheric NO_2 column (estimated using the 1-sigma standard deviation of the V_{strat}), σ_{Vftrop} is the uncertainty of the free troposphere NO_2 column (estimated using the 1-sigma standard deviation of the V_{ftrop}). R_{CV} is the GEM-MACH calculated surface VMR to PBL column ratio, and σ_R is the uncertainty of that ratio (estimated using the 1-sigma standard deviation of the R_{CV}). The means of R_{CV}, V_{Pan}, V_{strat}, and V_{ftrop} are used in the uncertainty estimation.

Besides the UP method, another simple approach to estimate uncertainty is to compare the data product with another high-quality (lower uncertainty) coincident data. For example, if we assume that the Pandora direct-sun total column NO_2 data can represent the true value, we can estimate the uncertainty of Pandora zenith-sky total column NO_2 by calculating the 1-sigma standard deviation of their difference (referred to here as the SDD method):

$$\sigma_{VCD_{ZS}} = \sigma(VCD_{DS} - VCD_{ZS}). \qquad (9)$$

Similarly, if we assume that the in situ surface NO_2 VMR can represent the true value, the uncertainty of Pandora zenith-sky-based surface NO_2 VMR can be given by:

$\sigma_{C_{Pan}} = \sigma(C_{insitu} - C_{pan})$. (10)

Also, if there is systematic bias between the two datasets, it can be removed and the random uncertainty can be calculated by:

$\sigma_{VCD_{ZS}} = \sigma(VCD_{DS} - k_1 VCD_{ZS})$, (11)

$\sigma_{C_{Pan}} = \sigma(C_{insitu} - k_2 C_{pan})$, (12)

where k_1 and k_2 are the slopes in the linear fits with intercept set to zero (e.g., slopes in Figs. 2 and 6). This method is referred to here as the unbiased SDD. These three uncertainty estimation methods (UP, SDD, and unbiased SDD) were all implemented, and the results are summarized in Table A1. The results show that Pandora zenith-sky total column NO_2 data have a 0.09-0.12 DU uncertainty that is about twice to the Pandora direct-sun total column nominal accuracy (0.05 DU, at 1-sigma level). When using the UP method, for the worst-case scenario, the Pandora zenith-sky total column NO_2 have a 0.17 DU uncertainty (i.e. using minimum of AMFs to estimate the upper limit of uncertainty). The estimated Pandora zenith-sky-based surface NO_2 VMR data have uncertainties from 4.8 to 6.5 ppbv. In Eqn. 8, the contributions of the V_{Pan}, V_{Strat}, V_{ftrop}, and R_{CV} terms to the total uncertainty are 36%, 2%, 0.3%, and 62 %, respectively. This result indicates that the uncertainty in the Pandora zenith-sky-based surface NO_2 VMR is dominated by the uncertainties of Pandora zenith-sky total column NO_2 and the modelled column-to-surface conversion ratio (R_{CV}). However, note that this uncertainty budget depends on the NO_2 vertical distributions, and hence may vary from site to site; e.g., in Toronto, tropospheric column NO_2 is typically 2-4 times higher than stratospheric column NO_2, and thus, the contribution to uncertainty from V_{Pan} is much larger than the corresponding contributions from V_{Strat} and V_{ftrop}. In addition, the uncertainty of Pandora direct-sun surface NO_2 VMR is also estimated and provided in Table 1. It shows slightly better results than for zenith-sky-based surface NO_2 VMR.

Table 1. Estimated uncertainties for Pandora zenith-sky total column and surface NO_2.

Estimation method	$\sigma_{VCD_{ZS}}$ (DU)	$\sigma_{C_{Pan-ZS}}$ (ppbv)	$\sigma_{C_{Pan-DS}}$ (ppbv)
UP	0.12	6.5	5.4
SDD	0.09	5.1	5.0
unbiased SDD	0.09	4.8	4.8

Section 3.1: The approach introduced for the zenith-sky AMF calculation is fully empirical and strongly inspired from the zenith-sky measurement mode used with Dobson and Brewer total ozone spectrophotometers. Basically the idea is to use simultaneous direct-sun and zenith-sky measurements to infer effective AMFs for the zenith-sky geometry. It is then assumed that the established relationship remains valid under moderately cloudy conditions (O4 is used to exclude thick diffusing clouds). I first note that the authors do not refer to the AMT publication by Tack et al. (2015) (https://www.atmosmeas- tech.net/8/2417/2015/) where a more physical approach to derive total and

tropospheric NO2 columns from zenith-sky measurements is described. Second, I see a major drawback in the empirical approach used here, which is that the total AMF for zenith-sky measurements is expected to be a strong function of not only the solar zenith angle but also the tropospheric column itself. In first approximation, one can assume that the stratospheric AMF will mostly follow the solar geometry (geometrical AMF) while the PBL AMF is approximately constant and close to one at any solar position. In consequence, for intermediate and low sun conditions, the stratospheric and PBL AMFs differ quite strongly and the total AMF depends on the relative amount of NO2 present in the PBL and in the stratosphere. I think that the classification could be improved substantially by taking this dependence into account (probably within an iterative scheme). The dependence on the season accounts somehow for this effect (since it implicitly accounts for the seasonality of the stratospheric NO2 column), but only in a very crude way.

The information from Tack et al. (2015) has now been included in Section 3.1 and this publication has been included in the References section:

In Tack et al. (2015), a more sophisticated four-step approach to derive total and tropospheric NO_2 columns from zenith-sky measurements was proposed, which involves using a RTM to calculate appropriate tropospheric AMFs. However, due to benefits from using the high-quality Pandora direct-sun total column NO_2 measurements, this work took a different but simple and robust approach to derive zenith-sky total column NO_2.

We thank the referee for their insightful suggestions about improving the empirical AMF calculations. The current empirical AMFs are limited to high and intermediate sun conditions (i.e., SZA< 75°). In the future, when deriving low-sun empirical ZS AMFs, we will try the suggested iterative scheme to account for the stronger influence of PBL NO_2. We think this suggestion is valuable, and we have now included this information in the manuscript (Appendix A) as follows:

In addition, the current empirical AMFs are limited to high and intermediate sun conditions (i.e., SZA< 75°). For low-sun conditions, the total AMF for zenith-sky measurements is expected to be a strong function of not only the SZA, but also the tropospheric column itself. Thus, for future work to derive low-sun empirical zenith-sky AMFs, the stronger influence of PBL NO_2 has to be accounted for (i.e., the geometry form AMFs are not enough).

Section 3.1, L. 19: " : : : VCD_Emp shows less SZA dependence than VCD_DS: : : " -> I think that VCD_NDACC is meant here instead of VCD_DS

Corrected.

In addition, VCD_{Emp} shows less SZA dependence than VCD_{NDACC} (see the increased bias for measurements made in larger SZA conditions in Figure 2b).

Section 3.1, last paragraph: note that the zenith-sky AMF could also be affected by aerosols present in PBL (together with NO2), due to their impact on the light path. Although the impact of aerosols is likely to be moderate, it certainly contribute to the uncertainty of the measurements and this should be mentioned.

Thank you for this comment. The following new text has been included in Section 3.1, last paragraph:

The derived zenith-sky total column NO_2 values are affected by both clouds and aerosols due to their impact on the light path. The presence of clouds and aerosols contributes to the uncertainty of the measurements. However, the impact of aerosols is expected to be moderate in most cases compared to that of clouds (e.g., Hendrick et al., 2011; Tack et al., 2015). Thus, this work has focused on evaluating the impact from clouds.

Section 3.2, L. 5-10: the finding that zenith-sky columns despite their larger uncertainties (in comparison to direct-sun data) show a better agreement with satellite measurements is quite surprising and interesting. I am not convinced by the argument stating that the air mass sampled by zenith-sky measurements is more representative of the air mass sampled by the satellite than the direct-sun. Considering the size of typical OMI pixels (approx. 20x20 km2), one can argue that both direct-sun and zenith-sky measurements are more local in nature, and therefore maybe another explanation can be found. Could it be that direct-sun and zenith-sky measurements sample different meteorological conditions (e.g. different wind patterns), or maybe that zenith-sky are generally more homogeneously distributed around the overpass time of the satellite so that the average value becomes more representative? Please comment on these issues.

We were not able to find a solid reason for this finding (zenith-sky measurements have better agreement with OMI). The measurement frequency was given in Section 2.1.1 and the coincidence criteria were provided in Section 3.2. We agree with the referee that both direct-sun and zenith-sky measurements are more local when compared with OMI measurements. However, with the coincidence criteria used (± 30 min), we are not sure that meteorological conditions are the key factor. Also, we did not average the ground-based data. The nearest (in time) measurement that was within ± 30 min of OMI overpass time was used (see Section 3.2). Therefore, it is not an averaging issue. In addition, when taking the standard error of the fitting and the confidence level of R into account, the difference between zenith-sky and direct-sun data is not significant (i.e., in Fig. 4 from panels a to d, the slopes with standard error are 0.64 ± 0.02, 0.67 ± 0.02, 0.70 ± 0.04, and 0.71 ± 0.03; the 95% confidence interval for R values are 0.45 to 0.63, 0.61 to 0.75, 0.43 to 0.77, and 0.60 to 0.86). In general, we think it is probably a case of coincident error, i.e., compared to Pandora DS, both OMI and Pandora ZS underestimate the local NO_2 at Toronto (see Section 3.2 and Figure 2). More discussion of this issue has now been included in Section 3.2:

The better correlation and lower bias for zenith-sky versus direct-sun might be a case of coincident error, i.e., compared to Pandora direct-sun, both OMI and Pandora zenith-sky total column NO_2 underestimate *the local NO_2 at Toronto (see Figure 2). When taking into account the standard error of the fitting and the confidence level of R, the difference between zenith-sky and direct-sun data is not significant (i.e, in Fig. 4 from panesl a to d, the slopes with standard error are 0.64 ± 0.02, 0.67 ± 0.02, 0.70 ± 0.04, and 0.71 ± 0.03; the 95% confidence intervals for R values are 0.45 to 0.63, 0.61 to 0.75, 0.43 to 0.77, and 0.60 to 0.86).*

Section 4.1: the method used to convert NO2 column measurements into surface concentrations implies a lot of approximations/assumptions. The uncertainties associated to these assumptions should be better described. Equation 3 starts from the total NO2. For the zenith-sky case, this column already contains quite a large uncertainty (cf. previous point). Then a correction for the stratospheric column and the free tropospheric column is made. The stratospheric column is taken from photochemically-corrected stratospheric NO2 OMI measurements without any further verification. Can we exclude any possible systematic bias between OMI and ground-based measurements? Was there any attempt to verify that OMI and Pandora measurements do agree well under clean conditions? Also monthly mean OMI data are used. Can we safely neglect the day-to-day variability in the stratospheric NO2 content (e.g. in Spring it is known that transport patterns can produce short term variations of the stratospheric NO2)?

There could be a systematic bias between OMI and ground-based data but most likely the bias found at Toronto is related to tropospheric NO_2. Thus, we decided to assess the OMI stratospheric NO_2 data and used it in the algorithm. In this work, we tested the algorithm using both OMI daily and monthly mean stratospheric NO_2, and we did not find any significant differences in the derived surface NO_2. The reason is that for Toronto, tropospheric NO_2 accounts for 73 % of the total column amounts on average (see Section 3.1) around local noon. During rush hours, the percentage of tropospheric NO_2 should be even higher. Also, when using OMI daily stratospheric NO_2, we need to interpolate the data for cloud days, which will also increase uncertainty. One of the reasons that we decided to use monthly mean OMI stratospheric NO_2 was also to reveal the robustness of the method (i.e., not highly dependent on satellite measurements). Some of this information has now been included in Appendix B:

Note that the strength of this bias is related to 1) the NO_2 profile (weights between stratospheric and tropospheric NO_2), and 2) the observation geometry (direct-sun or zenith-sky). In general, an urban site with direct-sun observation should have less impact from the stratospheric diurnal variation. On the other hand, a rural site with zenith-sky observation should have significant impact.

As regards the free-tropospheric NO_2 content, it is taken directly from model data. How large and uncertain is this contribution? Finally the last step is based on the assumption that the modelled column to surface concentration ratio is representative of the actual ratio. Is this assumption expected to be correct in all cases? What about the impact of the limited horizontal resolution of the model? In any case, I think that all these uncertainties are responsible for the large scatter of the correlations shown in Figure 6. As such they should be discussed with a little bit more of attention. This being said, I agree that the average behaviors found (and illustrated in Figs. 9-11) are quite convincing and demonstrate well the potential of the data set.

The current method is sophisticated in that outputs from several different chemical transport models (with different strengths) were included. Currently, we are working to provide estimate of uncertainties of the final data products (Pandora surface NO_2). Based on the Toronto datasets, the mean of free tropospheric NO_2 is 0.03 ± 0.01 DU (mean ± 1σ) (GEOS-Chem), while the boundary layer NO_2 is 0.37 ± 0.29 DU (GEM-MACH), and stratospheric NO_2 is 0.10 ± 0.02 DU (OMI monthly mean stratospheric NO_2, with Pratmo to account for diurnal variation). Thus, the uncertainty contributions from GEOS-Chem, OMI, and Pratmo are much less than the uncertainty contribution from GEM-MACH. An uncertainty estimation section is now included as Appendix D (see the reply to previous comments), which addresses the uncertainties associated with the model inputs.

The assumption that the modelled column to surface concentration ratio is representative of the actual ratio has challenges, especially for shallow boundary layer conditions (e.g., Figure 11). However, we think the model can better represent the real conditions for most cases (taking emissions, atmospheric dynamics, and chemistry into account), when comparing to other existing methods (which simply assume that NO_2 is uniformly mixed through the entire PBL). The current GEM-MACH operational forecast model has spatial resolution of 10 km × 10 km (information was provided in Section 2.2.1). Based on our current results, the model shows some bias compared to in situ data, which could have contributions from the spatial resolution.

Finally this study makes use of Pandora direct-sun completed by zenith-sky measurements. At no point in the paper, the potential of extending the data set with multi-axis measurements providing more information on the tropospheric NO2 is mentioned, although this would be a logical evolution for a follow up activity. Please consider adding this in the perspectives.

Thank you for this comment. Text about the potential of Pandora multi-axis measurements has been added to the Conclusions section.

Currently, the standard Pandora observation schedule includes direct-sun, zenith-sky, and multi-axis scanning measurements (i.e., measuring at multiple viewing angles). At present, multi-axis measurement algorithms are still under development, but in the future, by using the multi-axis measurements and optimal estimation techniques (e.g., Rodgers, 2000) or the five angles O2O2-ratio algorithm (Cede, 2019), it may be possible for Pandora measurements to be used to derive NO_2 *tropospheric profiles and columns.*

Reference

Bogumil, K., Orphal, J., Homann, T., Voigt, S., Spietz, P., Fleischmann, O. C., Vogel, A., Hartmann, M., Kromminga, H., Bovensmann, H., Frerick, J. and Burrows, J. P.: Measurements of molecular absorption spectra with the SCIAMACHY pre-flight model: instrument characterization and reference data for atmospheric remote-sensing in the 230-2380 nm region, J. Photochem. Photobiol. A, 157, 167–184, doi:10.1016/s1010-6030(03)00062-5, 2003.

Cede, A.: Manual for Blick Software Suite 1.6, 2019.

Chance, K. V. and Spurr, R. J. D.: Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum, Appl. Opt., 36, 5224–5230, doi:10.1364/AO.36.005224, 1997.

Hendrick, F., Pommereau, J. P., Goutail, F., Evans, R. D., Ionov, D., Pazmino, A., Kyrö, E., Held, G., Eriksen, P., Dorokhov, V., Gil, M. and Van Roozendael, M.: NDACC/SAOZ UV-visible total ozone measurements: improved retrieval and comparison with correlative ground-based and satellite observations, Atmos. Chem. Phys., 11, 5975–5995, doi:10.5194/acp-11-5975-2011, 2011.

Hermans, C., Vandaele, A. C., Fally, S., Carleer, M., Colin, R., Coquart, B., Jenouvrier, A. and Merienne, M.-F.: Absorption cross-section of the collision-induced bands of oxygen from the UV to the NIR, in Weakly Interacting Molecular Pairs: Unconventional Absorbers of Radiation in the Atmosphere, edited by C. Camy-Peyret and A. A. Vigasin, pp. 193–202, Springer, Germany., 2003.

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

Rothman, L. S., Jacquemart, D., Barbe, A., Chris Benner, D., Birk, M., Brown, L. R., Carleer, M. R., Chackerian, J. C., Chance, K., Coudert, L. H., Dana, V., Devi, V. M., Flaud, J. M., Gamache, R. R., Goldman, A., Hartmann, J. M., Jucks, K. W., Maki, A. G., Mandin, J. Y., Massie, S. T., Orphal, J., Perrin, A., Rinsland, C. P., Smith, M. A. H., Tennyson, J., Tolchenov, R. N., Toth, R. A., Vander Auwera, J., Varanasi, P. and Wagner, G.: The HITRAN 2004 molecular spectroscopic database, J. Quant. Spectrosc. Radiat. Transfer, 96, 139–204, doi:10.1016/j.jqsrt.2004.10.008, 2005.

Tack, F., Hendrick, F., Goutail, F., Fayt, C., Merlaud, A., Pinardi, G., Hermans, C., Pommereau, J.-P. and Van Roozendael, M.: Tropospheric nitrogen dioxide column retrieval from ground-based zenith–sky DOAS observations, Atmos. Meas. Tech., 8(6), 2417–2435, doi:10.5194/amt-8-2417-2015, 2015.

Vandaele, A. C., Hermans, C., Simon, P. C., Carleer, M., Colin, R., Fally, S., Mérienne, M. F., Jenouvrier, A. and Coquart, B.: Measurements of the NO_2 absorption cross-section from 42 000 cm^{-1} to 10 000 cm^{-1} (238 -1000 nm) at 220 K and 294 K, J. Quant. Spectrosc. Radiat. Transfer, 59, 171–184, doi:10.1016/s0022-4073(97)00168-4, 1998.