



LuftBlick report 2021001

Quality Assurance for Earth Observation - QA4EO

Final Report WP2126 "PGN Products Quantitative Uncertainty"

	Name	Company	Date
prepared by	Martin Tiefengraber	LuftBlick	3 Dec 2021
	Karin Kreher	LuftBlick	3 Dec 2021
checked by	Alexander Cede	LuftBlick	3 Dec 2021

Table of Content

[Summary and Conclusions](#)

[2 Introduction](#)

[2.1 Acronyms and Abbreviations](#)

[2.2 Applicable Documents](#)

[2.3 Reference Documents](#)

[3 Quantification of the uncertainty when using an extraterrestrial reference for NO₂ column retrieval](#)

[3.1 Usability evaluation of extraterrestrial reference spectra and consequences on the fitting setup selection](#)

[3.2 Estimation of the common uncertainty by comparisons to the operational NO₂ dataset](#)

[4 Quantification of the algorithm error of the operational PGN NO₂ product](#)

[4.1 Spectral fitting \(L2Fit\) algorithm](#)

[4.1.1 Impact of NO₂ column load and assumed effective temperature](#)

[4.1.2 Impact of O₃ column load and spectral feature](#)

[4.1.3 Impact of AOD and Angström \$\alpha\$](#)

[4.2 Direct sun AMF \(L2\) algorithm](#)

[4.2.1 Impact of effective NO₂ layer height](#)

[5 Propagation of uncertainty within the PGN data analysis](#)

[5.1 Overview](#)

[5.2 L0 Uncertainty](#)

[5.3 L1 Uncertainty](#)

[5.4 L2Fit Uncertainty](#)

[5.5 L2 Uncertainty](#)

[6 Appendix](#)

[Altzomoni](#)

[Brussels-Uccle](#)

[Wakkerstroom](#)

[HoustonTX](#)

[Yokosuka](#)

[Athens](#)

[MexicoCity](#)

[Helsinki](#)

[Bangkok](#)

[Ulsan](#)

Document Change Record

Issue	Page	Date	Observations
1	All	2021-12-03	First version

Summary and Conclusions

The WP2126 “PGN products quantitative uncertainty” of the ESA project “Quality Assurance for Earth Observation” (QA4EO) [AD1] had essentially three goals. A brief summary and conclusion are given here for each of them.

Goal 1) Model the common uncertainty related to NO₂ column retrievals applying an extraterrestrial reference spectrum

This is discussed in section 3. The idea was to investigate if the operationally applied PGN NO₂ retrieval, using a measured reference spectrum, could be accompanied by an retrieval using an extraterrestrial reference spectrum which would have e.g. the advantage no field calibration would be needed. However, using an extraterrestrial reference for the NO₂ retrieval leads to additional uncertainties due to unaddressed differences between the instrument which measured the reference spectrum and the instrument making the actual trace gas measurements. To quantify this common uncertainty component, retrievals made with the extraterrestrial reference spectrum are compared with the operational retrievals made with the measured reference spectrum.

This comparison was undertaken for 11 PGN stations and details of each of the comparisons is provided in the Appendix. An unexpected, but rather consistent NO₂ slant column difference between both retrievals for all 11 stations could be identified, with a median value of -90.2×10^{-6} mol/m². This overall bias is most likely attributed to residual NO₂ absorption structures in the self-claimed “extraterrestrial” data sources. The 1-sigma standard deviation of 68.1×10^{-6} mol/m² over all stations is finally considered as the common uncertainty related to NO₂ retrievals with an extraterrestrial spectrum.

Goal 2) Quantify the algorithm error of the operational PGN NO₂ product

For the determination of the total uncertainty of a data product, it is vital to understand and quantify the intrinsic errors connected to the applied algorithm. Using simulated data, we quantify the impact of selected relevant factors on the operational direct sun PGN NO₂ product in section 4, with the following conclusions:

1. The basic algorithm error for total column NO₂ is vanishingly low (< 0.05 %).
2. The O₃, NO₂ column load and the variation in aerosol parameters (AOD, Angström α) have a negligible effect on the NO₂ retrieval error (< 0.3 %).
3. Potentially unaccounted spectral features (e.g. caused by etaloning or by alignment changes between optical parts) can lead to a quite substantial AMF dependent bias in the order of 5 % (highest at smallest AMFs).

Goal 3) Modify the Blick processing software (BlickP) to propagate the uncertainty from the different sources into the Pandora L1, L2Fit and L2 data

As part of this project, BlickP V1.8 was extensively restructured and updated to take into account the correct nomenclature and usage of data uncertainties as recommended by NPL. Three types of uncertainties (independent, common and structured) are distinguished, which differ from each other by the correlation length of the associated uncertainty along a certain "dimension". Here, this dimension is wavelength for L1 data data (i.e. the measured spectra) and time for L2Fit data (i.e. the fitted slant columns) and L2 data (i.e. the total columns). The uncertainties from the different sources are then propagated into the Pandora L1, L2Fit and L2 data. This is described in detail in section 5.

As part of goal 2 also a correlative uncertainty study led by NPL was proposed, but has not been finished as of this date. LuftBlick could not deliver a needed forward model in a shareable format so far and is still working on this task.

2 Introduction

This is the final report of WP2126 “PGN Products Quantitative Uncertainty” of the ESA project “Quality Assurance for Earth Observation” (QA4EO) [AD1] for the time period 1 November 2020 to 1 November 2021. The report is structured into the following sections:

- Section 1 - Summary and Conclusions
- Section 2 - Introduction
- Section 3 - Quantification of the uncertainty when using an extraterrestrial reference for NO₂ column retrieval
- Section 4 - Quantification of the algorithm error of the operational PGN NO₂ product
- Section 5 - Propagation of uncertainty within the PGN data analysis

2.1 Acronyms and Abbreviations

AD	Applicable Document
AMF	Air Mass Factor
AOD	Aerosol Optical Depth
AtmVar	Atmospheric Variability
BlickSFA	Bick spectral fitting algorithm
BlickP	Pandora Processing Software
DQF	Data quality flag
DU	Dobson Units
L0 data	Level 0 data
L1 data	Level 1 data
L2 data	Level 2 data
NO ₂	Nitrogen Dioxide
O ₃	Ozone
PGN	Pandonia Global Network
QA4EO	Quality Assurance Framework for Earth Observation
RD	Reference Document
SC	Slant Column
SUSIM	Solar Ultraviolet Spectral Irradiance Monitor
SZA	Solar Zenith Angle
UARS	Upper Atmosphere Research Satellite
U _i	Independent Uncertainty
U _c	Common Uncertainty
U _s	Structured Uncertainty

2.2 Applicable Documents

[AD1] Quality Assurance for Earth Observation project [Annex B, Statement of Work for LuftBlick], SERCO Contract QA4EO/SER/SUB/04, 2020.

2.3 Reference Documents

- [RD1] Robert L Kurucz. New atlases for solar flux, irradiance, central intensity, and limb intensity. *Memorie della Societa Astronomica Italiana Supplementi*, 8:189, 2005.
- [RD2] Gröbner, J., Kröger, I., Egli, L., Hülsen, G., Riechelmann, S., and Sperfeld, P.: The high-resolution extraterrestrial solar spectrum (QASUMEFTS) determined from ground-based solar irradiance measurements, *Atmos. Meas. Tech.*, 10, 3375–3383, <https://doi.org/10.5194/amt-10-3375-2017>, 2017.
- [RD3] Michael E VanHoosier. Solar ultraviolet spectral irradiance data with increased wavelength and irradiance accuracy. In *SPIE's 1996 International Symposium on Optical Science, Engineering, and Instrumentation*, pages 57–64. International Society for Optics and Photonics, 1996.
- [RD4] Jonathan Mittaz, Christopher J Merchant, and Emma R Woolliams. Applying principles of metrology to historical earth observations from satellites. *Metrologia*, 56(3):032002, 2019.
- [RD5] Cede A., Manual for Blick Software Suite 1.8, [Blick Software Suite Manual Version 1.8](#).
- [RD6] Timmer, J., & Koenig, M., On generating power law noise. *Astronomy and Astrophysics*, 300, 707, 1995.

3 Quantification of the uncertainty when using an extraterrestrial reference for NO₂ column retrieval

The operational NO₂ total column product within PGN applies a measured reference spectrum (SyntRef) in the retrieval algorithm (this product is abbreviated NO2SyntRef). Consequently, in order to retrieve absolute values, the slant column (SC) amount and temperature of NO₂ in this measured spectrum needs to be quantified. A reliable calibration of this amount (done using a Langley like approach) demands a suitable amount of clear sky data, usually collectable within a couple of weeks.

Retrieving NO₂ columns from applying an extraterrestrial reference spectrum from the literature (ExtRef), would, per definition, provide absolute values right away (this product is abbreviated NO2ExtRef). This gives rise to two advantages:

- (A) In theory, absolute NO₂ columns can be retrieved without further calibration and could bridge the waiting time for the operational NO₂ product to be calibrated.
- (B) Routine comparison to the operational NO₂ product helps to identify potential instrumental changes (more details below).

However, applying ExtRef in the retrieval is affected by enhanced uncertainties due to unaddressed differences between the instrument which measured the reference spectrum and the instrument taking the measurements. Since this uncertainty affects each SC retrieval equally (infinite correlation length in time), it can be considered as common uncertainty (compare chapter 5.1).

To quantify this uncertainty component, we compare retrievals of NO2ExtRef with NO2SyntRef for 11 stations, spanning a wide range of NO₂ pollution scenarios (e.g. remote, suburban, urban). By considering NO2SyntRef datasets as the reference, we estimate the common uncertainty of NO2ExtRef from the median difference to NO2SyntRef. This is shown in detail in section 3.2. Before that, we want to highlight notable aspects regarding the usability of ExtRef in section 3.1.

3.1 Usability evaluation of extraterrestrial reference spectra and consequences on the fitting setup selection

As an initial approach, we retrieve NO₂ExtRef with the same fitting setup as used for NO₂SyntRef (400 to 470 nm, 4th order closure polynomial). We tested two ExtRefs, both composed of the high resolution information of Kurucz [RD1] and a radiometric correction with (for the wavelengths of interest)

- SUSIM/Atlas 3 [RD3] (< 426 nm) and Gröbner [RD2] or (now referred to as SUSIM)
- only Gröbner (now referred to as Gröbner)

For more details about the compilation of the ExtRefs, please refer to the Blick manual [RD5].

These ExtRef retrievals, together with the operational NO₂SyntRef, have been exemplarily tested at two collocated Pandora instruments in Rome (Pandora 117s1, 138s1) for the 5th of September 2020. The retrieved time series of NO₂ is shown in Figure 3.1.1, where the left (right) panel shows results for Pandora 117 (138): NO₂SyntRef is considered the reference and is shown in green.

Two observations have to be highlighted here:

- Both NO₂ products based on ExtRef (SUSIM in red and Gröbner in blue) show a clear SC bias to the reference dataset. This bias can be removed, if a certain SC amount is added to the retrieved SCs (amount is written in parenthesis in the figure legend).
- This SC bias of NO₂ExtRef to NO₂SyntRef is different for both
 - the selected ExtRef source and
 - the Pandoras.

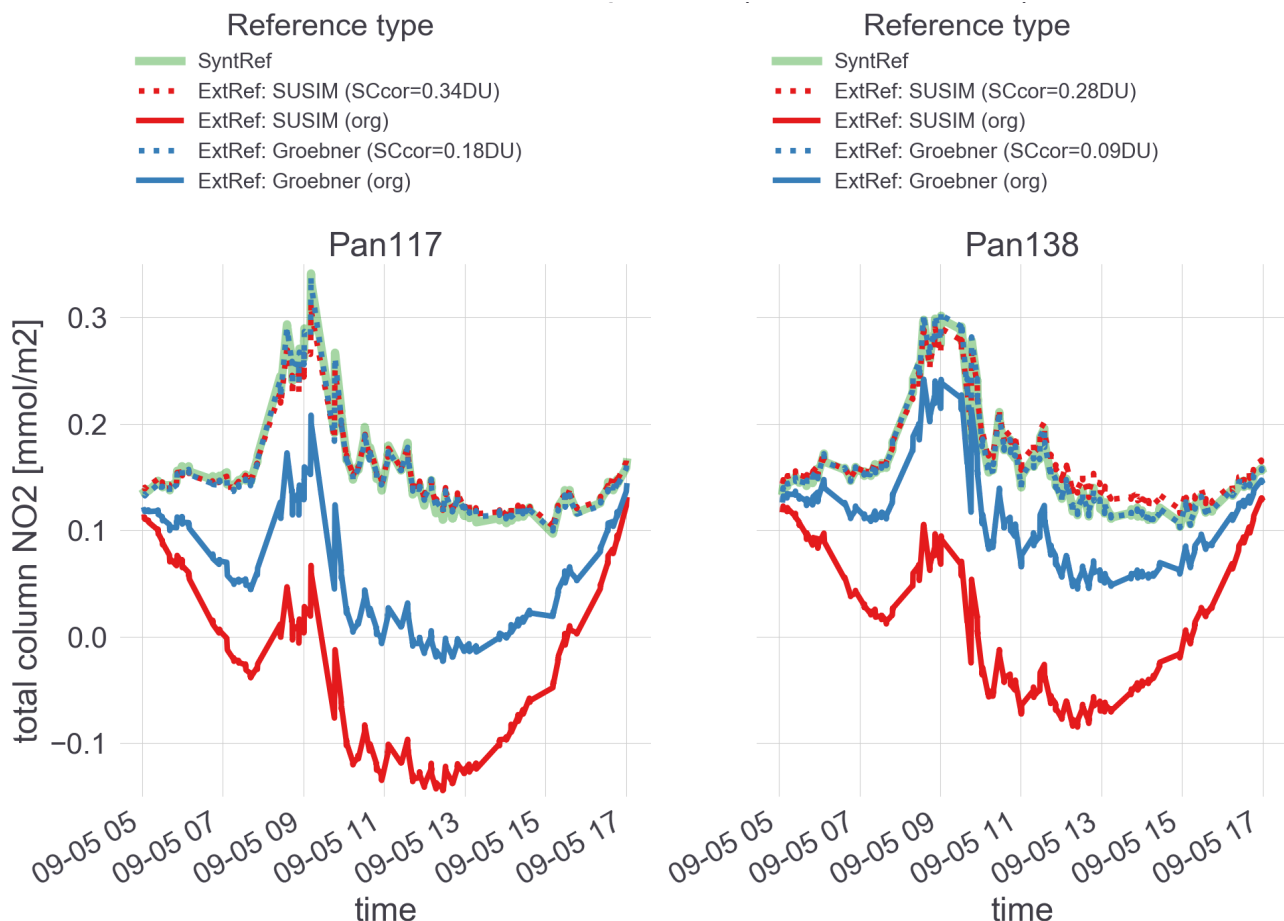


Figure 3.1.1: Comparison of total column NO₂ retrievals for different reference types for two Pandoras (left #117, right #138) at Rome Sapienza University (September 5th, 2020). The reference datasets (using SyntRef) are shown in green and two datasets using different ExtRef versions are shown in red (Kurucz+SUSIM) and blue (Kurucz+Gröbner). An intrinsic SC bias (listed in the figure legend) is evident in the retrievals using ExtRef.

If we would correct this SC bias, a remarkable agreement to the reference dataset can be seen. This raises confidence in the general usability of NO2ExtRef.

But next we want to address observation (I). A similar pattern, namely a clear SZA/AMF dependent diurnal variation of total NO₂, usually indicates a calibration error (meaning the SC amount in the reference spectrum hasn't been guessed correctly). The magnitude of the apparent SC "corrections" we see here, however, even indicates that the reference spectrum hasn't been corrected for NO₂ absorption at all. This is of course counterintuitive considering the label "extraterrestrial".

The Kurucz reference dataset is based on measurements taken at the Kitt Peak observatory in Arizona, USA (published 1984) with additional telluric spectrum corrections applied in 2005. However, according to the publication, NO₂ was indeed not considered in these corrections and is hence still present in the spectral fine structure.

The dataset measured by SUSIM (Solar Ultraviolet Spectral Irradiance Monitor) as payload of UARS (Upper Atmosphere Research Satellite) was an extraterrestrial mission and therefore does not contain any telluric features. Hence the radiometric correction using this spectrum is expected to not introduce any NO₂ features.

The Gröbner reference dataset was compiled from measurements taken at the atmospheric observatory in Izana, Tenerife, in September 2016. Also, according to the publication, no NO₂ absorption was considered in the correction of telluric spectral features. Consequently we can expect that in the course of the radiometric correction of Kurucz, additional NO₂ features are introduced.

These insights allow us therefore to understand part (A) of observation (II). Both the fine structure of Kurucz and the Gröbner spectrum are expected to contain residual NO₂ features. This would explain why a SC correction is needed and why for the Gröbner version (no straightforward connection of Kurucz and Gröbner) the bias is different. Please note that different fitting windows and closure polynomials have been tested which did not change this pattern. The final determination of the SC bias in the ExtRef is detailed in the next section.

One additional comment about the fitting window for NO2ExtRef: we have to limit the starting wavelength to > 430 nm. This was necessary due to artificial spectral features which have been identified in the Kurucz ExtRef. Figure 3.1.2 shows the daily average spectral fitting residuals for the same dataset as before, where the reference dataset (NO2SyntRef) is given in green and the NO2ExtRefs in red (SUSIM) and blue (Gröbner). The artificial spectral structures of SUSIM below 426 nm are evident. Although this observation suggests preferring the Gröbner ExtRef over the SUSIM for this spectral region, we were not really considering this for the moment (the whole PGN database makes use of the ExtRef of SUSIM and would need to be recalibrated). The feature might also be related to the fact, that 426 nm is the wavelength, where what we call "ExtRef: SUSIM" transitions to the Gröbner spectrum as indicated at the beginning of section 3.1.

More insights are provided by the multi-station comparison between NO2ExtRef and NO2SyntRef. This comparison is shown next and allows to

- Note that knowing the residual NO_2 content in ExtRef, would allow to produce an indeed NO_2 absorption free version of ExtRef.

To model the common uncertainty of NO2ExtRef, retrievals from 11 PGN stations have been analysed by investigating the NO₂ slant column differences using a SyntRef (operational PGN approach) and an ExtRef (based on Kurucz+SUSIM). Table 3.2.1 shows an overview of the NO2 time series available at the 11 PGN stations used for this study.

Station	Latitude	Longitude	Station classification	Pandora number	Start date	End date	NO ₂ SC mean & std	Number of pairs
---------	----------	-----------	------------------------	----------------	------------	----------	-------------------------------	-----------------

							in 10 ⁻⁶ mol/m ²	
Bangkok	13.7847	100.5400	urban	190s1	20/05/2021	10/10/2021	-173.27 +/-7.36	8465
Brussels - Uccle	50.7980	4.3580	suburban	162s1	17/03/2020 31/08/2020	28/05/2020 30/06/2021	-130.22 +/-11.91	62337
Wakkerstroom	-27.3493	30.1438	suburban	159s1	20/12/2019	10/10/2021	-56.11 +/-17.87	133966
MexicoCity -Vallejo	19.4830	-99.1470	urban	157s1	24/10/2019	09/10/2021	-210.70 +/-15.43	204508
Ulsan	35.5745	129.1896	urban	150s1	11/09/2021	10/10/2021	-83.78 +/-7.37	4852
Yokosuka	35.3207	139.6508	urban	146s1	29/02/2020 29/10/2020	26/07/2020 06/10/2021	1.11 +/-8.83	89335
Athens - NOA	37.9878	23.7750	urban	119s1	08/06/2016 16/12/2020	01/04/2020 10/10/2021	4.23 +/- 63.71	190415
Helsinki	60.2037	24.9612	suburban	105s1	18/06/2021	10/10/2021	-91.46 +/- 6.25	54975
Altzomoni	19.1187	-98.6552	remote	65s1	23/01/2019 26/01/2021	09/03/2020 10/10/2021	-87.39 +/-10.36	44065
HoustonTX	29.7200	-95.3400	urban	25s1	08/03/2021	10/10/2021	-129.61 +/- 9.50	62272
Busan	35.2353	129.0825	urban	20s1	03/03/2021	10/10/2021	-105.50 +/- 5.01	49

Table 3.2.1: List of 11 PGN stations, their latitude, longitude and station classification, the Pandora instrument number, and start and end date of the NO₂ time series. The last two columns list the mean difference and standard deviation between the two NO₂ slant column data sets, and the amount of data points available for each comparison.

The mean value and standard deviation for each PGN station comparison is summarised in the 2nd to last column of table 3.2.1 with the mean difference values ranging from approximately zero (or to be precise 1.11×10^{-6} mol/m²) for Yokosuka in Japan to -210.7×10^{-6} mol/m² for MexicoCity-Vallejo in Mexico. Figures showing the NO₂ slant column time series retrieved with each of the two retrieval algorithms (NO₂SyntRef and NO₂ExtRef) and the NO₂ slant column differences between the two retrievals are included in the Appendix for all 11 stations.

Figure 3.2.1 provides a summary of all investigated time series, showing a box plot for the 11 stations which is sorted by the size of the median of each of the NO₂ slant column difference data sets. The overall median value for these 11 stations is found at Altzomoni with -90.23×10^{-6} with a 1 sigma standard deviation of $\pm 68.11 \times 10^{-6}$ mol/m² over the median values of each of the 11 stations.

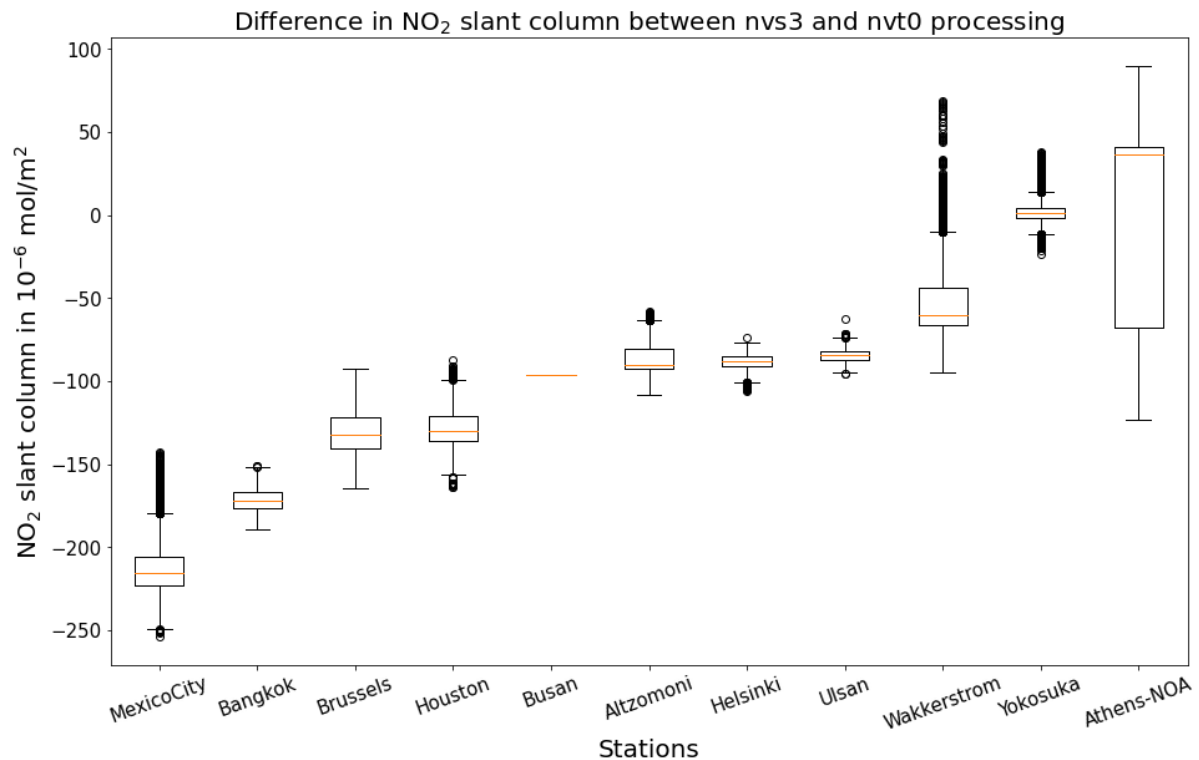


Figure 3.2.1: Overview of all 11 timeseries. The box plot shows the median of the dataset for each station (horizontal orange line in the middle of box), as well as the interquartile ranges (the ends of the boxes) and the minimum and maximum values of the chosen datasets (the far end of the “whiskers”). Outliers are plotted with an open circle. Only data pairs with NO₂ AMF < 2 are included.

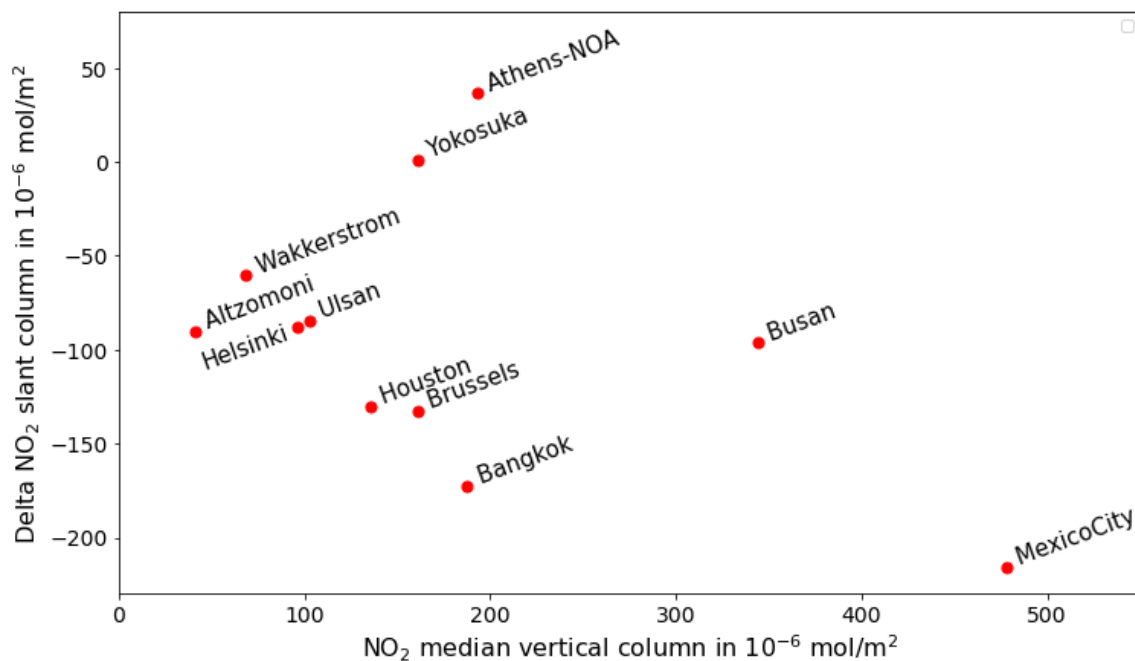


Figure 3.2.2: The median NO₂ slant column difference between using NO2ExtRef and NO2SyntRef is plotted against the median NO₂ vertical column value for each of the stations.

No clear pattern could be detected yet regarding what causes the size difference in the NO_2 slant column difference between retrievals. To investigate further if there is a relationship between more polluted locations and the size of the NO_2 SC difference, Figure 3.2.2 shows the median value for each of the 11 time series (also shown in Figure 3.2.1) plotted against the corresponding median NO_2 vertical column for each of the stations. Again, no clear pattern between remote, suburban and urban stations (see Table 3.2.1) can be observed.

Since uncertainty-exceeding calibration errors related to $\text{NO}_2\text{SyntRef}$ have been cross-checked again and hence can be excluded, the most likely driver for the varying differences is changes in the instrumental characteristics. Let's assume a change in the protocolled absolute instrument sensitivity ("absolute calibration") due instrument transport and setup. If a SyntRef is taken from the measurement site (=standard case), this feature cancels out in the retrieval. This is however not the case for ExtRef , since ExtRef still "relies" on a still sensitivity determined in the lab. This additional spectral structure is expected to slightly change cross-correlations of the fitting parameters, resulting in a constant SC bias in the retrieval (off actually all fitting parameters). This is also shortly addressed in the next section. From the examples in the appendix, rather stable instruments like the one in Wakkerstroom or more unstable instruments like the one in Athens can be identified. These more unstable instruments drive the common uncertainty qualification.

If we recap the two advantages mentioned in the introduction of this section, we can say that

ad (A): Absolute values using ExtRef are only available either if the detected SC bias is added to the retrievals or a new, NO_2 ExtRef is produced.

ad (B): A routine comparison of $\text{NO}_2\text{SyntRef}$ and NO_2ExtRef would indeed help to indicate instrumental changes.

4 Quantification of the algorithm error of the operational PGN NO_2 product

An important aspect in the estimation of the total uncertainty of a data product is the quantification of the intrinsic errors connected to the applied algorithm. It, at the same time, constitutes the accuracy limit or minimum uncertainty of the product.

Since this cannot be reliably done based on measured data, simulated data need to be modelled and applied to the algorithm. While it is in principle straightforward to simulate direct sun spectral irradiances, it is of particular importance to consider instrumental features like the full slit function (including spectral stray light), detector electronics (e.g. non-linearity) and sensitivities adequately in the simulation. This has been done for this study by incorporating the instrumental features from a regular Pandora (here Pandora 30) in the simulations.

Next to instrumental features, a standard atmospheric setup is the basis for the following simulations: Rayleigh scattering, aerosol extinction (Angstrom approach) and gas absorption from O_3 , NO_2 , H_2O and O_2O_2 at standard profiles (and US standard temperature profile) was considered.

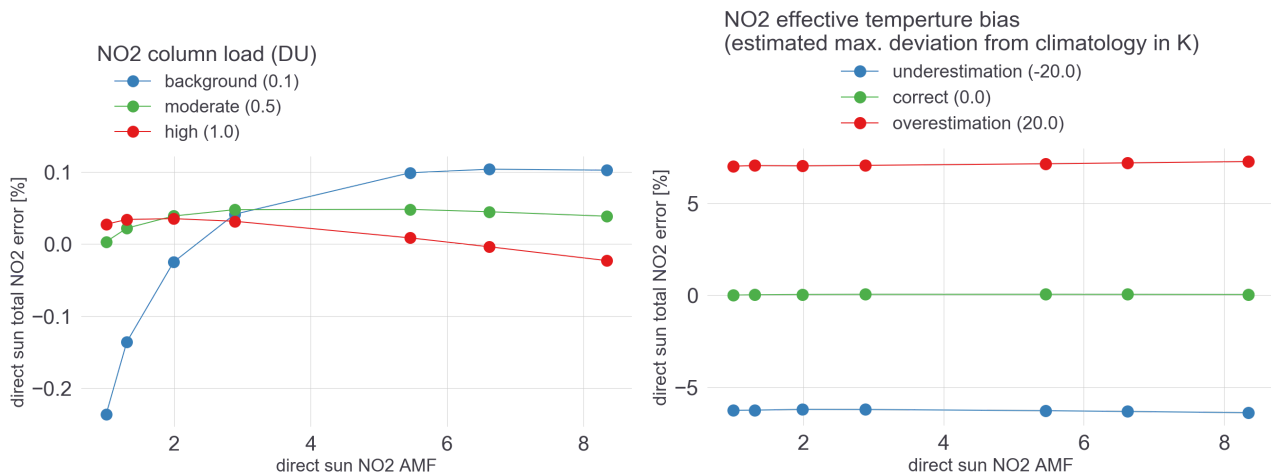
In the following we quantify a selection of factors impacting the algorithm error for the operational PGN direct sun NO_2 product. We picked factors of interest and where, to our knowledge, the impact is notable. Please note that some factors are just shown for interest and are not considered algorithm error, because they are already explicitly budgeted in (those factors are identified in the text). In section 4.1 we start out with factors affecting the spectral fitting algorithm and look at factors impacting the AMF calculation in section 4.2.

4.1 Spectral fitting (L2Fit) algorithm

4.1.1 Impact of NO₂ column load and assumed effective temperature

The NO₂ column load (left figure panel) has a negligible effect on the NO₂ retrieval error. Still negligible but slightly higher are NO₂ retrievals at background (=stratospheric) NO₂ loads at smaller AMFs.

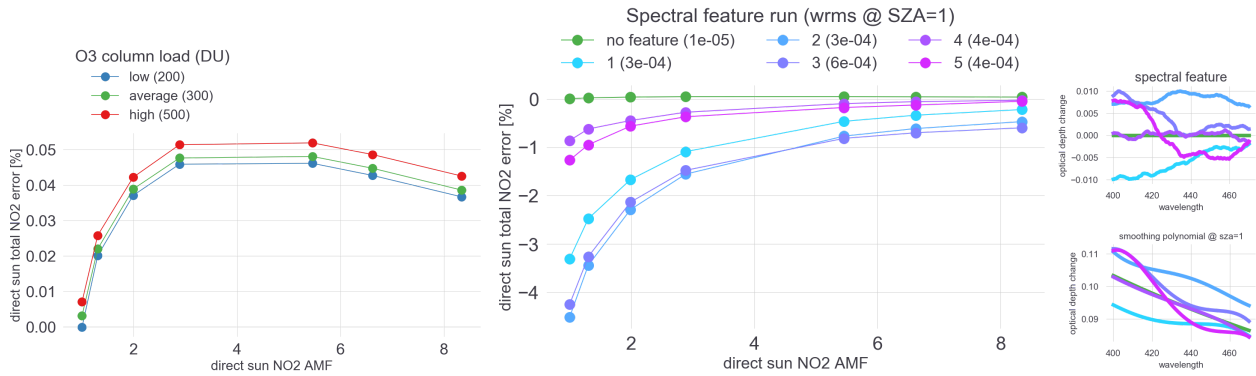
A notable AMF invariant bias can be expected, if the effective temperature for the NO₂ cross section is estimated wrongly (right figure panel). Note, this uncertainty component is already factored in as part of the structured uncertainty and is just shown for completeness. In this example a boundary layer temperature is suggested, where a possible over/underestimation of ± 20 K is assumed. The consequence would be a NO₂ retrieval error of $\sim \pm 6.5$ % (note, a wrong temperature estimation in the spectral fitting indeed is a relative error), where an underestimation in temperature leads to an underestimation in NO₂ columns.



4.1.2 Impact of O₃ column load and spectral feature

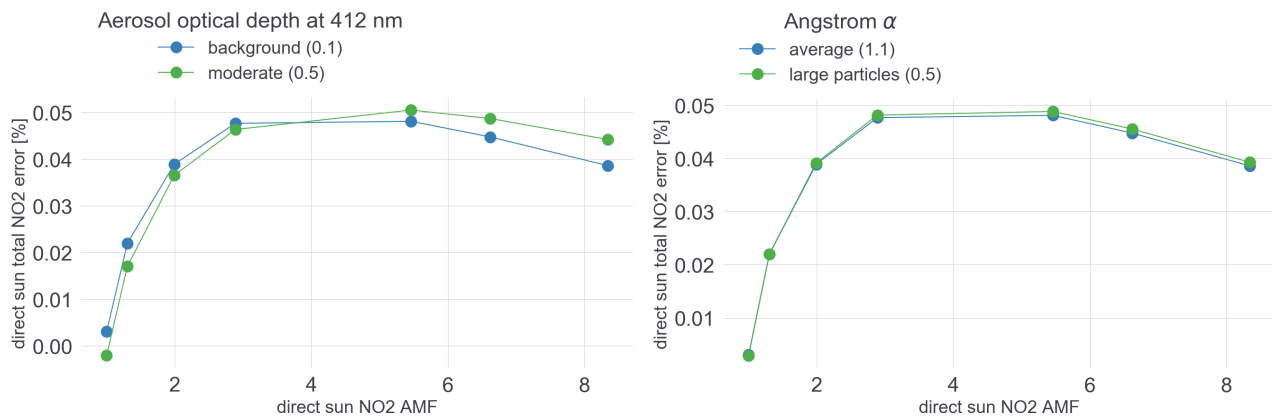
While the spectral fitting algorithm for NO₂ is very robust against changes in O₃ column amounts (left figure panel), a quite substantial bias can be expected when unaccounted spectral features disturb the fitting (right figure panel). Such features could emerge e.g. from etaloning or stimulated by alignment changes between optical parts.

The shown example represents a moderately strong (compare wrms values in the figure legend), constant feature, which is based on a power law noise ("colored" noise) simulation [RD6]. The corresponding spectral features (smoothing polynomials) are depicted in the embedded plot on the top (bottom) right. The impact is strongest for small AMFs (leveling out for higher AMFs due to the AMF division) and hence introduces (in this case) negative SC biases.



4.1.3 Impact of AOD and Angström α

Variation in either the AOD (left figure panel) or Angström α are totally negligible. The smooth spectral changes stimulated by changes in aerosol parameters, are easily absorbed by the spectral fitting closure polynomials.

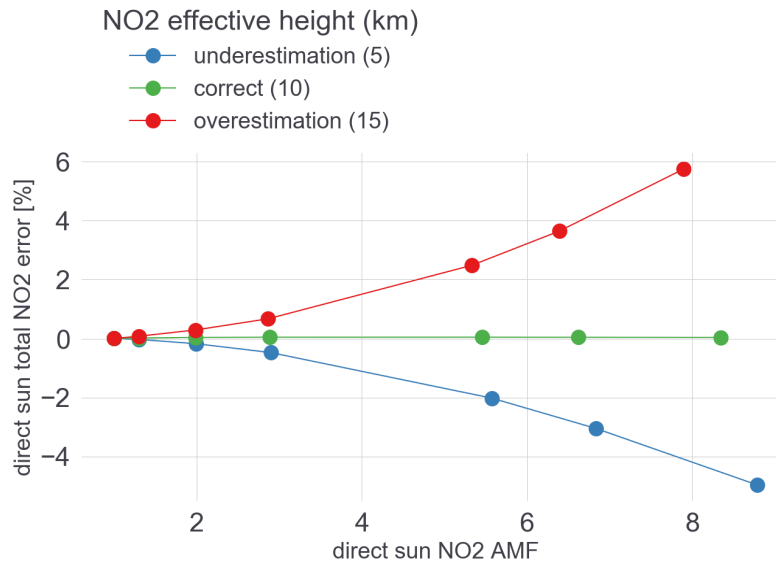


4.2 Direct sun AMF (L2) algorithm

4.2.1 Impact of effective NO₂ layer height

The effective NO₂ layer height is basically the only significant variable in the calculation of the direct sun AMF and impacts mainly high AMFs. Also here, the related uncertainty is already budgeted in and is shown for completeness only.

Since the PGN NO₂ direct sun algorithm calculates the layer height from estimating the stratospheric-tropospheric split (by using the given stratospheric NO₂ climatology and adjust it based on the retrieved SC from two retrievals using a stratospheric and boundary layer effective temperature), the layer height under/overestimation should be well within ± 5 km. The resulting NO₂ retrieval error is below 1 % for AMFs $\sim < 4$ and reaches up to ~ 6 % for high AMFs, generally leading to NO₂ column overestimation when the effective height is overestimated.



Overall we observe a very small basic algorithm error ($< 0.05\%$), with negligible additional effects of NO_2 and O_3 loads, as well as different aerosol conditions on the retrieval quality of total NO_2 (all below 0.3%).

As mentioned above, uncertainties related to cross-section temperature and effective height estimation are not considered algorithm errors and are already part of the uncertainty budget.

A notable impact could be revealed from spectral features which potentially disturb the measurement system. Since these features are unknown by nature, we can just assume their structure and magnitude in the simulation by following what we know from experience. Based on that, SC biases in the order of 5% are thinkable and would therefore introduce an artificial AMF dependent diurnal variation.

5 Propagation of uncertainty within the PGN data analysis

5.1 Overview

The main output products of the Blick processing software (BlickP) are spectra (L1), slant columns (L2Fit) and total and tropospheric column amounts or profiles or surface concentrations (L2). Most of these data come with associated uncertainties. As part of this project, BlickP was modified to propagate the uncertainty from the different sources into the Pandora L1 data (i.e. the measured spectra), the Pandora L2Fit data (i.e. the fitted slant columns) and the Pandora L2 data (i.e. the total columns).

For the naming and meaning of the uncertainties the Blick Software Suite follows the guidelines laid out by Mittazet al. [RD4]. Three types of uncertainties are distinguished, which differ from each other by the correlation length of the associated uncertainty along a certain "dimension". In the Blick Software Suite, this dimension is wavelength for L1 data and time for L2Fit and L2 data.

1. **Independent uncertainty:** the correlation length along the dimension for the independent error is zero.

An example for L1 data is the read noise in a certain pixel (i.e. at a certain wavelength), which is totally uncorrelated to the read noise in any other pixel (wavelength). An example for L2Fit data is the photon noise propagated into the slant column amount measured at a certain moment, which is totally

uncorrelated to the propagated photon noise for measurements taken at any other time. The uncertainty associated with an independent error is called "Independent uncertainty" and is symbolized with U_I .

2. **Common uncertainty:** the correlation length along the dimension for the common error is infinite.

An example for L1 data is a bias in the radiometric calibration due to a faulty current for the calibration lamp in the laboratory. While this bias affects smaller wavelengths more than higher wavelengths, there is nevertheless full correlation between the error in one wavelength and the error in any other wavelength. An example for L2Fit data is an error in the assumed slant column in the reference spectrum. This error affects all retrieved slant columns using the same reference spectrum in the same way, hence the error at a certain measurement is fully correlated to the error for measurements taken at any other time. The uncertainty associated with a common error is called "Common uncertainty" and is symbolized with U_C .

3. **Structured uncertainty:** the correlation length along the dimension for the structured error is larger than zero, but not infinite.

An example for L1 data is an error in the flat field correction around a certain wavelength region. Such an error affects all the pixels inside this wavelength region in the same way but is uncorrelated to pixels at a different part of the spectrum. An example for L2Fit data is a difference between the effective temperature of a trace gas used in the spectral fitting (assuming that the temperature is NOT fitted itself) and the true effective temperature of this gas in the atmosphere. This introduces an error in the retrieved slant column, which is highly correlated to measurements taken around the same time, but in general not correlated to measurements taken at times farther away. E.g. if an effective ozone temperature of 225 K is used in the spectral fitting, but the true effective ozone temperature is 228 K at 10:00 in the morning of 27 October, this causes approximately a -1% error in the retrieved ozone slant column. The next measurement on this day at 10:02 will still suffer almost exactly the same error, since the true temperature has hardly changed in the 2 minutes. However, a few days later, on 3 November, the true temperature has in general changed and might be 225 K, which means the error due a mismatch of the effective temperature is then 0 and not correlated to the error from 27 October at 10:00. The uncertainty associated with a structured error is called "Structured uncertainty" and is symbolized with U_S .

For the total uncertainty U of a single (L1, L2Fit or L2) data point, we simply combine U_I , U_C and U_S as shown in equation 6.1:

$$U = \sqrt{U_I^2 + U_C^2 + U_S^2}$$

When the data are averaged, e.g. by building the mean spectrum over a certain wavelength range, or the mean column amount over a certain time interval, the combined uncertainty associated with the mean is a combination of the individual $U_I(i)$, $U_C(i)$ and $U_S(i)$. $i=1$ to n is the index for a single data point out of the n data points averaged. Here we look at the two "extreme" cases.

In the first situation, the structured errors are fully correlated along the dimension. In this "short" case, the total uncertainty of the mean value, called $U(n, \text{short})$, is given by:

$$U(n, \text{short}) = \frac{1}{n} \cdot \sqrt{\sum_{i=1}^n [U_I(i)^2] + \left[\sum_{i=1}^n U_C(i) \right]^2 + \left[\sum_{i=1}^n U_S(i) \right]^2}$$

Hence the independent uncertainty of the mean is "reduced" compared to the individual values, but the common and structured uncertainties are not. An example for this would be the mean column amount over a rather short time period, e.g. 10 min, in which we assume the data with respect to mismatch of the true and assumed effective trace gas temperature to be fully correlated.

The other extreme case assumes the structured uncertainties to be uncorrelated along the dimension. In this "long" case, the total uncertainty $U(n, \text{long})$, is given by:

$$U(n, \text{long}) = \frac{1}{n} \cdot \sqrt{\sum_{i=1}^n [U_I(i)^2] + \left[\sum_{i=1}^n U_C(i) \right]^2 + \sum_{i=1}^n [U_S(i)^2]}$$

Here, the structured uncertainty "behaves" like the independent uncertainty. An example for this would be the mean column amount over a long time period, e.g. one year, when we assume that the temperature used in the spectral fitting is from a climatology that represents very well the average true effective temperature over this year. Then we could say that the temperature errors are a mixture of over- and underestimations and can therefore be approximated as uncorrelated overall.

It is important to note that not all possible uncertainty sources are included in the Blick Software Suite at this moment and therefore even the total uncertainty is still incomplete and likely to be underestimated. Therefore the output products also include data quality flags (DQF), which are described in the BlickP user manual [RD5] in detail.

Any uncertainty sources which have currently not yet been taken into account will be added to a later versions of the software. More details on the uncertainty outputs of the Blick Software Suite, and which uncertainty sources are (not) included are given in sections 5.2 - 5.5 and in the algorithm descriptions.

5.2 L0 Uncertainty

Since in most cases a measurement set is comprised of several individual spectrometer readings, i.e. the "number of cycles" is above 1, the L0 data contain the "Uncertainty of raw counts for each pixel divided by the square root of the number of cycles", $U_M(L0i)$, in addition to the "Mean over all cycles of raw counts for each pixel" (see table 18 in the Blick Software Suite Manual [RD5]). $U_M(L0i)$ does not represent one single error type as introduced in section 5.1, since it is in general a mixture of types. Therefore we use index "M" for "Measured".

Mathematically, $U_M(L0i)$ is simply the standard deviation over the individual spectra in case L0-output "Uncertainty indicator" equals 1, or the rms to a straight line fitted into the individual spectra in case the "Uncertainty indicator" equals 2. In both cases, it is also divided by the square root of the number of cycles.

Physically, $U_M(L0i)$ is a combination of the independent instrumental uncertainty (i.e. read noise and photon noise), possible other instrumental uncertainty (maybe caused by faulty pointing during part of the measurement), and the variability of the input radiation over the duration of the measurement set. For a light source, which is

considered stable, like a calibration lamp, the Blick Software Suite uses uncertainty indicator 1. Solar radiation reaching the Earth's surface over a short time interval (typically <1 min) is considered to have a linear intensity change with time in case the atmospheric transmission is constant, simply because the solar zenith angle is changing. Therefore the Blick Software Suite uses uncertainty indicator 2 for field measurements.

Hence when the variability of the input radiation is small and there are only insignificant systematic instrumental errors in the measurement, $U_M(L0_i)$ is mostly an independent instrumental uncertainty. The more input variability is present, the more $U_M(L0_i)$ includes a common component, since this input change will affect all pixels in a correlated way. For field measurements, we call the variability of the input radiation "Atmospheric Variability" (AtmVar), which is defined in the next section.

5.3 L1 Uncertainty

In the current version of the Blick Software Suite, there are two L1 outputs related to uncertainty. The "Independent instrumental uncertainty of L1 data for each pixel", $U_I(L1_i)$, and the "Atmospheric variability of L1 data for each pixel [%]", $AtmVar(L1_i)$.

$U_I(L1_i)$ is first determined after the L1 correction step "Dark correction" is applied. It is calculated based on equations discussed in the Blick Software Suite Manual [RD5], and shown below in a slightly different way:

$$U_I(L1_i) = \sqrt{\left(\frac{1}{n_{DC} \cdot n_{PIX}} + \frac{1}{n_{BC}}\right) \cdot \sigma_{DCi}^2 + \frac{1}{n_{BC}} \cdot GAIN \cdot L1_i}$$

The $L1_i$ are the corrected counts at pixel i , GAIN is ICF-entry "Gain [counts per electron]" and the dark variance σ_{DCi}^2 is given from the $L0$ data, i.e. $U_M(L0_i)$ for the dark measurements. n_{DC} and n_{BC} are the number of cycles for dark and bright counts respectively. n_{PIX} is set to 1 for the "Immediate dark method" and is the number of pixels on the detector for the "Dark map method" (see also Blick Software Suite Manual, section 6.4.2 and equation 61 and 73 [RD5]). Then $U_I(L1_i)$ is propagated through all other correction steps such as corrections for non-linearity, latency, flat field, conversion to count rates, temperature, stray light, sensitivity and wavelength.

$AtmVar(L1_i)$ is defined by the following equation:

$$AtmVar(L1_i) = \left[1 - \frac{U_I(L1_i)^2}{U_M(L1_i)^2}\right] \cdot 100$$

$U_M(L1_i)$ is initialized with $U_M(L0_i)$ as defined in section 5.2 and then propagated through all correction steps listed in the paragraph above just as it is done for $U_I(L1_i)$. $AtmVar(L1_i)$ is a percentage <100% indicating the magnitude of the atmospheric variability during the measurement interval. Small numbers mean the atmosphere was very stable. In practice the numbers can also be negative. Higher numbers mean there was more and more variability in the atmosphere. Numbers close to 100% are typically obtained, when a cloud moves in or out of the beam during a direct sun measurement.

There are special situations, where the L1 uncertainty output is not exactly as described in this section. This happens e.g. if only one cycle is measured and therefore no $U_M(L0_i)$ is given for the dark measurements. In that

case, the uncertainty is calculated in a different way, e.g. calculating the dark variance σ^2_{DCi} as shown in the Blick Software Suite Manual, equation 6, based on ICF-entry "Dark variance power fit coefficients" [RD5]. The different cases are indicated by L1 output "Indicator for uncertainty and atmospheric variability (see manual for exact meaning)" and described in table 36 of the Blick Software Suite Manual [RD5]. In the vast majority of situations, the indicator will have the value 10, which means there were more than one cycle for both the bright and dark measurements and the description given in this section applies.

There is no common or structured uncertainty given for L1 in the current version of the Blick Software Suite. The reason is that at this moment, there are no uncertainties associated with the different calibration parameters given in the ICF.

5.4 L2Fit Uncertainty

The L2Fit data include independent, $U_i(L2Fit)$, common, $U_c(L2Fit)$, and structured, $U_s(L2Fit)$, uncertainties for the following parameters retrieved by the Blick spectral fitting algorithm (BlickSFA) [RD5]:

- Slant column amounts for each fitted gas
- Effective temperatures for each fitted gas
- Ring spectrum pseudo-slant column in case it was fitted
- Each coefficient of the smoothing polynomial used in the fitting
- Each coefficient of the offset polynomial used in the fitting
- Each coefficient of the wavelength change polynomial used in the fitting
- Each coefficient of the resolution change polynomial used in the fitting

Additionally, the L2Fit output also gives a so-called "rms-based uncertainty", denoted $U_{rms}(L2Fit)$, for each of the parameters listed above. The detailed mathematics for the retrieval of all uncertainties is given in the Blick Software Suite manual, section 6.5.10 [RD5]. An overview of the error sources affecting L2Fit and whether they are included in the BlickSFA is given in table 5.1.

Table 5.1: Uncertainty output in L2Fit data. Column "Output" lists that L2Fit uncertainty output, in which the respective error source is taken into account. If this column is empty, then this source is not included in the BlickSFA yet.

Error source	Remark	Output
L1 data	$U_i(L1)$ is propagated into L2Fit and contributes to $U_i(L2Fit)$. Since no $U_c(L1)$ or $U_s(L1)$ is given in this software version, the L1 data do not contribute to $U_c(L2Fit)$ or $U_s(L2Fit)$ yet.	$U_i(L2Fit)$
Cross sections	Neither the choice of the cross sections nor their uncertainty is taken into account	
Algorithm deficiencies	The BlickSFA is an approximation and "suffers" from intrinsic deficiencies. Those are mostly caused by the use of pre-convoluted	

	parameters, e.g. the cross sections, and cross-correlation among the fitted gases. These effects are not considered in the current software version.	
Reference "Ext*"	No uncertainty is associated with this reference. This means that any systematic issues, arising from using a reference not measured by the instrument itself, are neglected.	
Reference "Synt*"	ICF-entry "Independent instrumental uncertainty of synthetic reference spectrum for FUNCFILT" contributes to $U_i(L2Fit)$, and the uncertainties in ICF-entry "Slant columns in synthetic reference spectrum for FUNCFILT" contribute to $U_c(L2Fit)$. Potential errors from a possible "mismatch" of the reference, e.g. when using a solar reference for lunar data, are not considered. Any error in ICF-entry "Wavelengths for synthetic reference spectrum for FUNCFILT [nm]" is not considered either.	$U_i(L2Fit)$ $U_c(L2Fit)$
Reference "Meas*"	As for uncertainties associated with L1 data, only the $U_i(L1)$ from the selected reference $U_i(L2Fit)$ measurement is included and contributes to $U_i(L2Fit)$.	
Reference "Ref_*"	If the external reference includes independent and/or common uncertainties, then they are propagated and contribute to $U_i(L2Fit)$ and/or $U_c(L2Fit)$.	$U_i(L2Fit)$ $U_c(L2Fit)$
Effective temperature	If the effective temperature of a gas is NOT fitted, then the given uncertainty from FSE "Gas temps" is propagated through the BlickSFA and contributes to output $U_s(L2Fit)$. Hence if FSE "Gas temps" contains totally unreasonable values, then unreasonable values for $U_s(L2Fit)$ are produced too.	$U_s(L2Fit)$
Molecular scattering	If FSE "Mol scatt" is NOT set to "NO", then the Blick Software Suite takes an estimation for the uncertainty of the surface pressure from a climatology, propagates it through the BlickSFA, and adds it to output $U_s(L2Fit)$.	$U_s(L2Fit)$
"Known" slant OD	If one or more heritage f-codes are used in the spectral fitting, then the output from those calculations is subtracted as "known" slant OD in the BlickSFA, the respective uncertainties are propagated and added to the final L2Fit output uncertainties.	$U_i(L2Fit)$ $U_s(L2Fit)$ $U_c(L2Fit)$

5.5 L2 Uncertainty

L2 data based on the "L2 Direct Algorithm" (see Blick Software Suite, section 6.6 [RD5]), i.e. total vertical column amounts and effective temperatures for each output gas, include independent, $U_i(L2)$, common, $U_c(L2)$, structured, $U_s(L2)$, total, $U(L2)$, and "rms-based", $U_{rms}(L2)$, uncertainties. The Blick Software Suite propagates the uncertainties coming from the L2Fit output and also adds uncertainty raising from the estimation of the direct AMF. Figure 5.5.1 shows as an example the different uncertainty contributions for total vertical NO_2 columns of Pandora 67 for a period of 15 min.

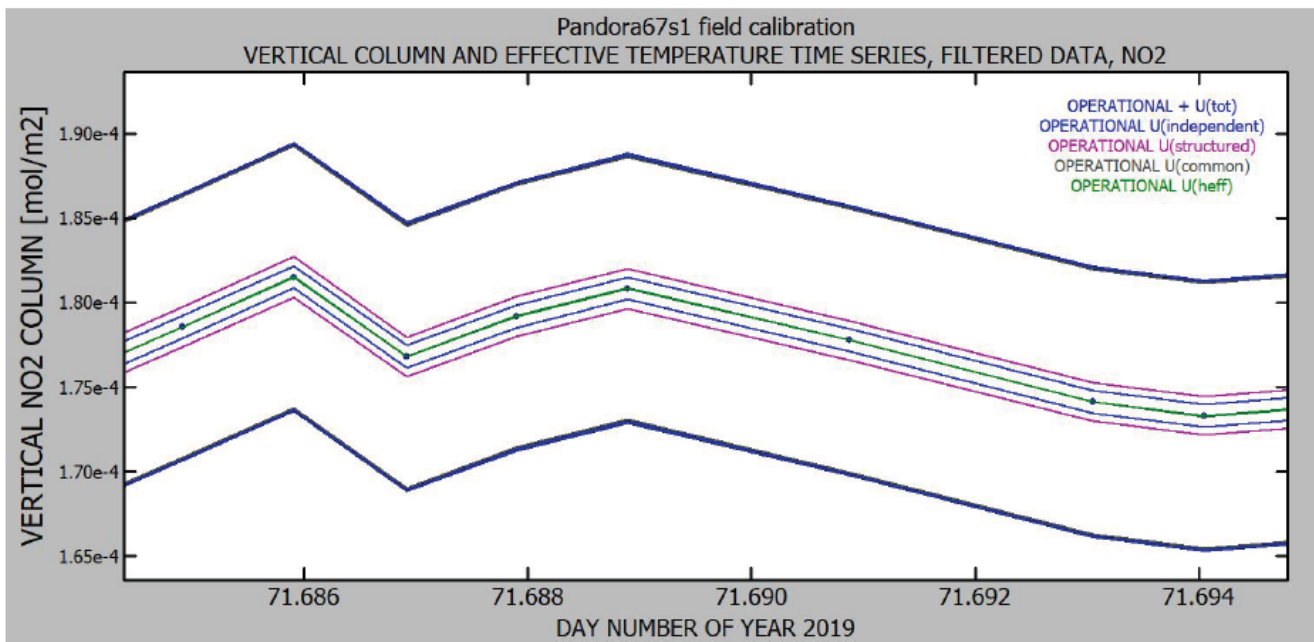


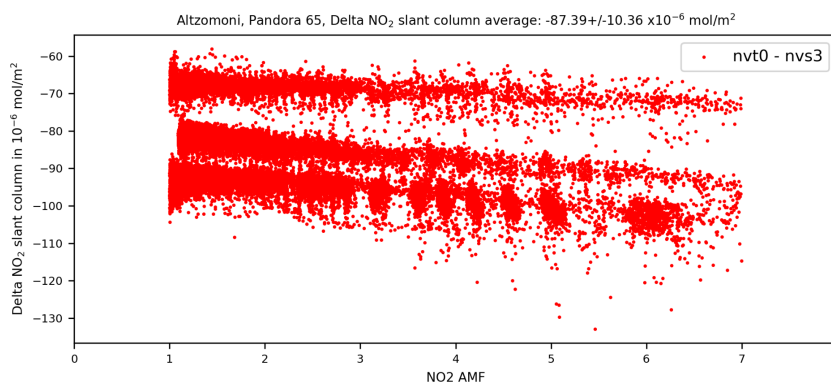
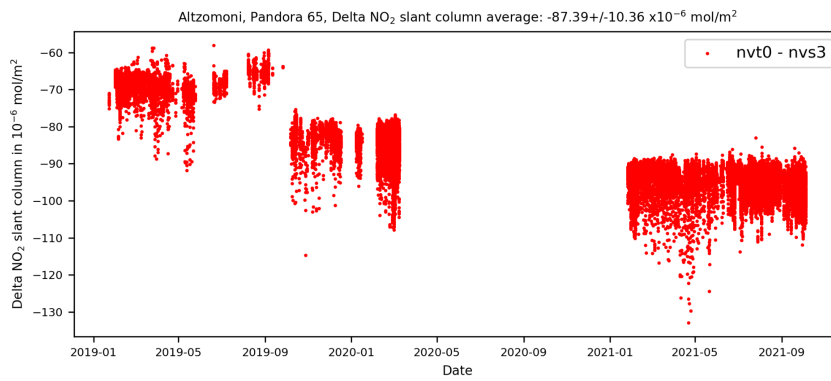
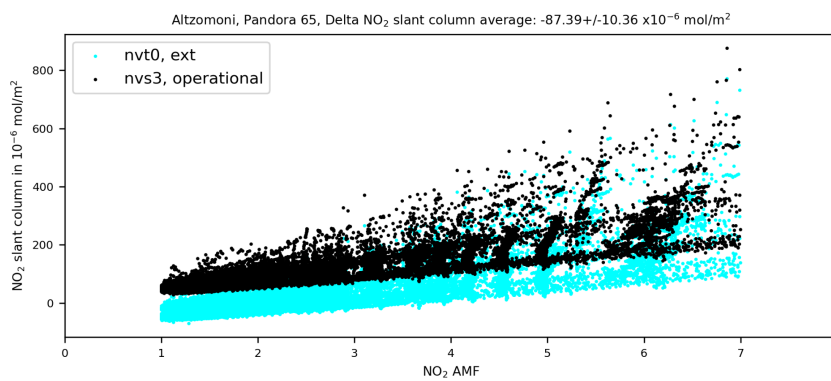
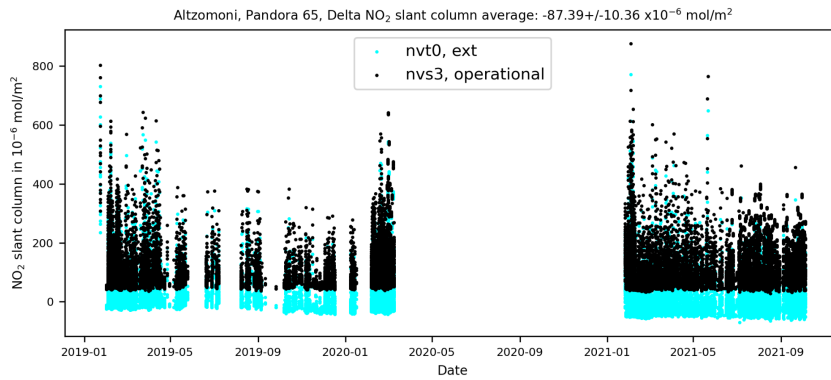
Figure 5.5.1: Vertical NO₂ column amount from Pandora 67 for a period of about 15 min on 12 Mar 2019 (dots) and different uncertainties as lines: AMF-uncertainty (green), independent uncertainty (thin dark blue), structured uncertainty (purple), common uncertainty (grey, mostly hidden under the dark blue line) and total uncertainty (thick dark blue).

For the tropospheric columns and surface concentrations obtained by the Air-Ratio Sky Algorithm, the Blick Software Suite produces an independent uncertainty output by simply propagating the independent uncertainties from the L2Fit data through equations listed in the Blick Software Suite Manual, section 6.7 [RD5]. No other (common or structured) uncertainties are produced yet in the software version. For the profile data, no uncertainty information is given yet.

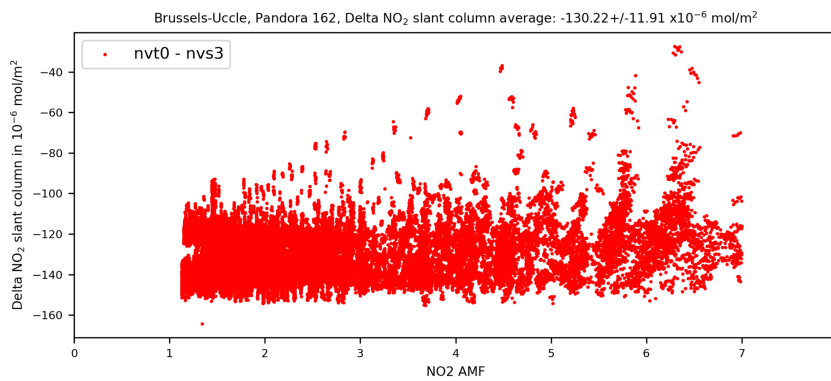
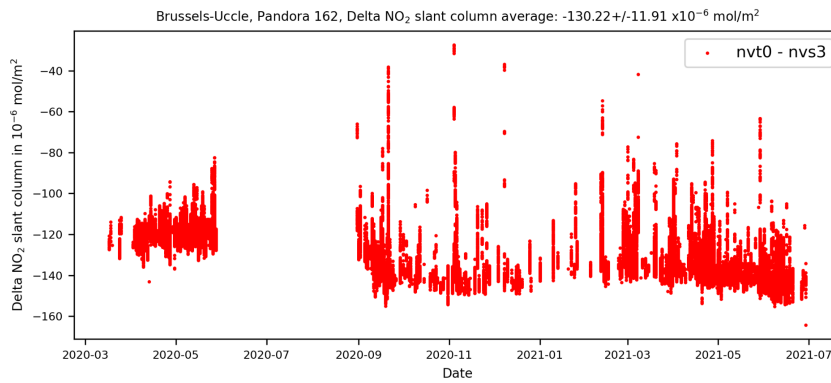
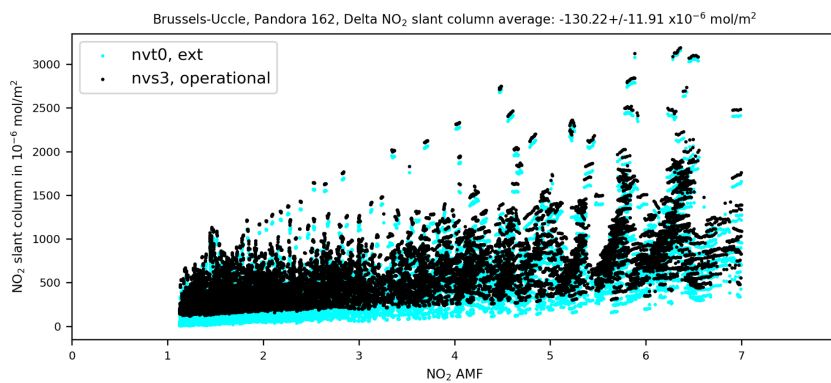
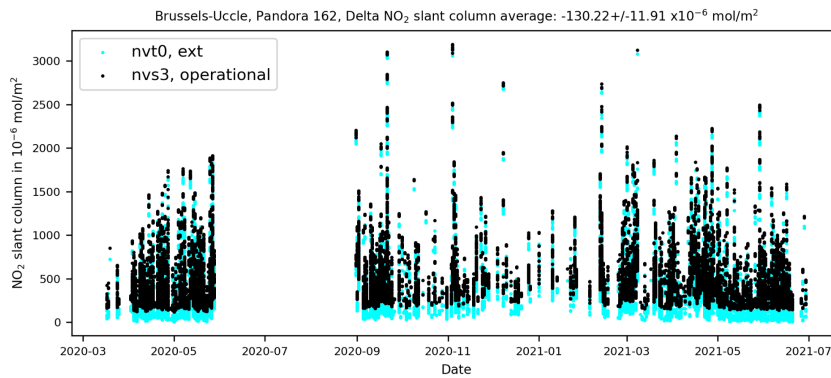
6 Appendix

The following figures show four plots for each PGN station used in the study discussed in section 3.2. The two top plots show the NO₂ slant column time series for both retrieval algorithms plotted against date and against the NO₂ airmass factor. The operational nvs3 algorithm uses a NO₂SyntRef (cyan) and the nvt0 algorithm uses a NO₂ExtRef (black). The two bottom plots show the difference between both NO₂ slant column time series plotted against date and airmass factor.

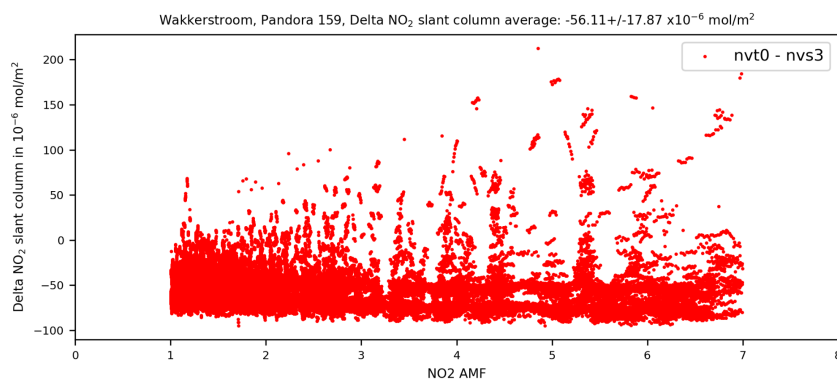
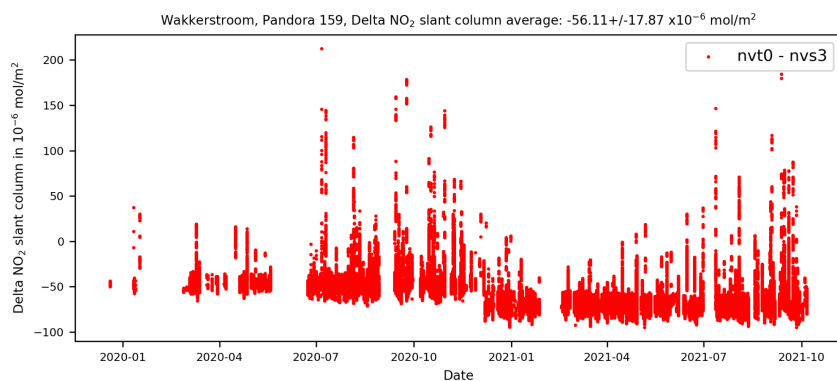
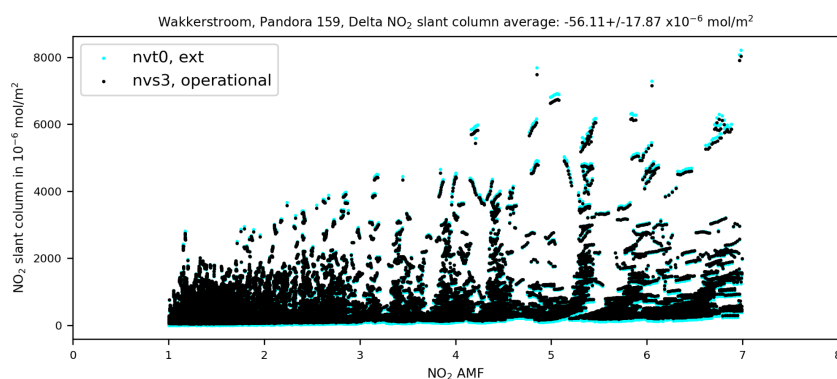
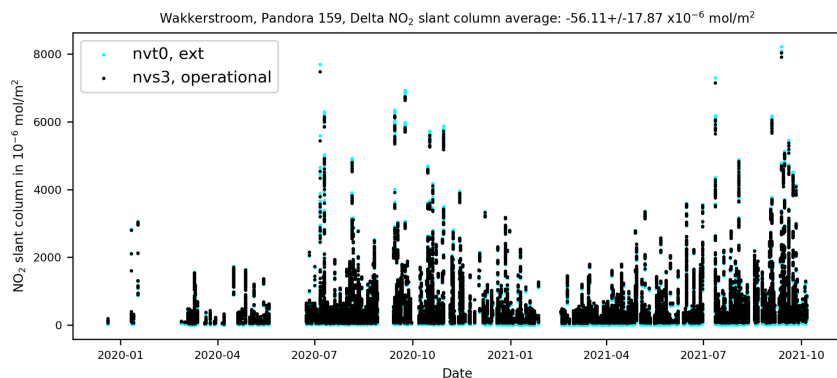
Altzomoni



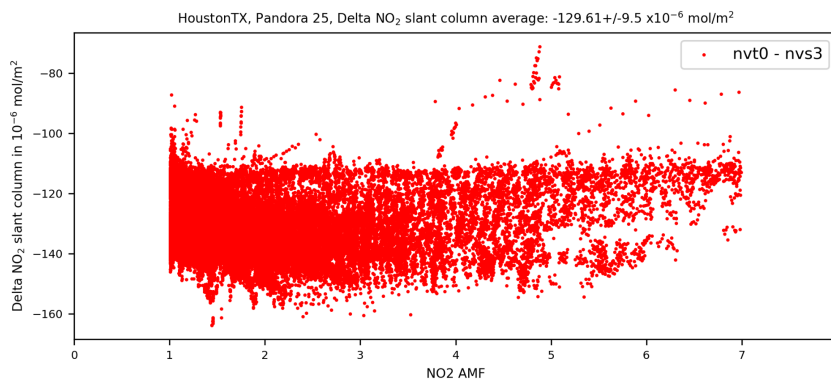
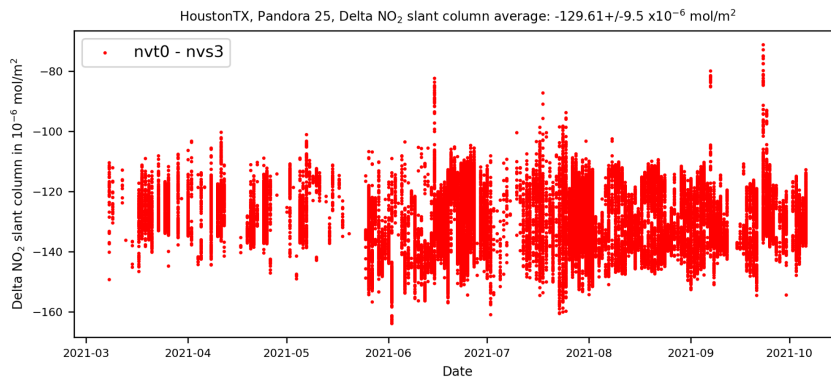
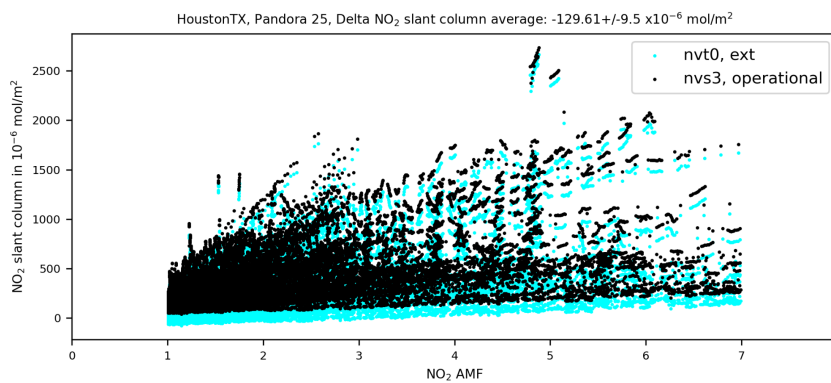
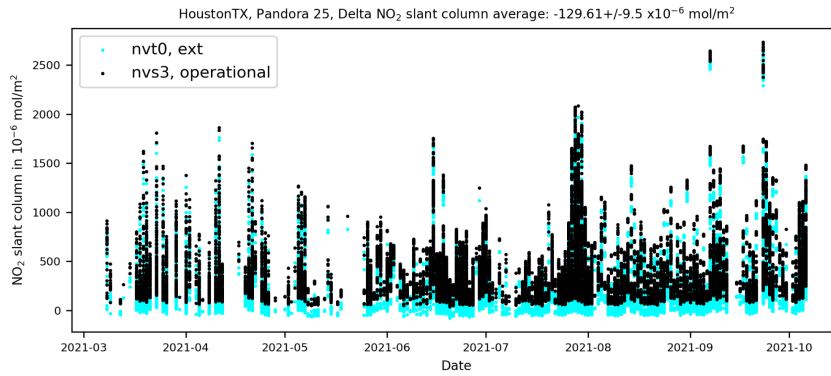
Brussels-Uccle



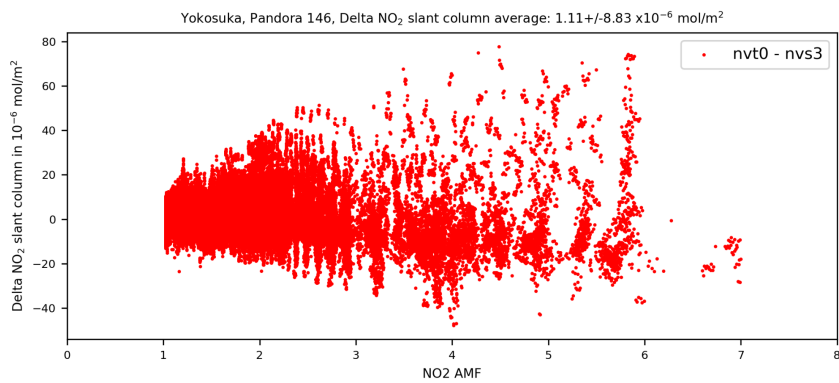
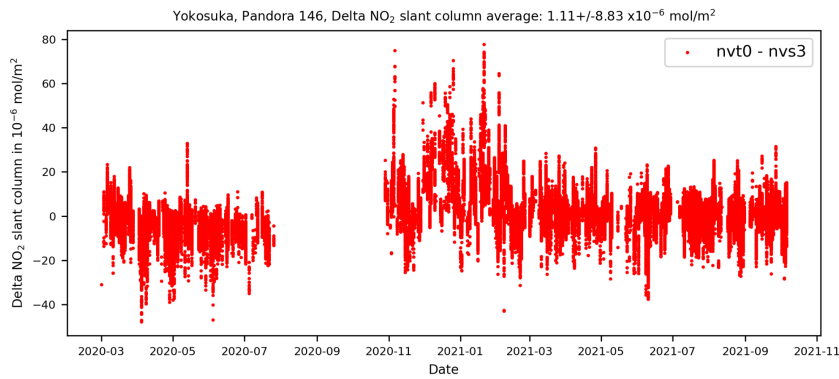
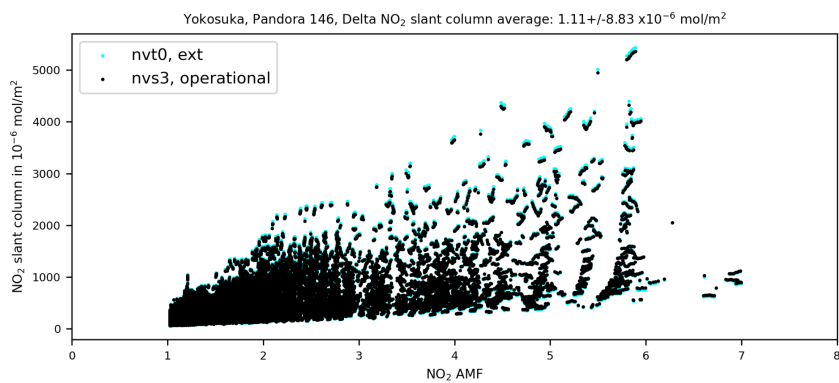
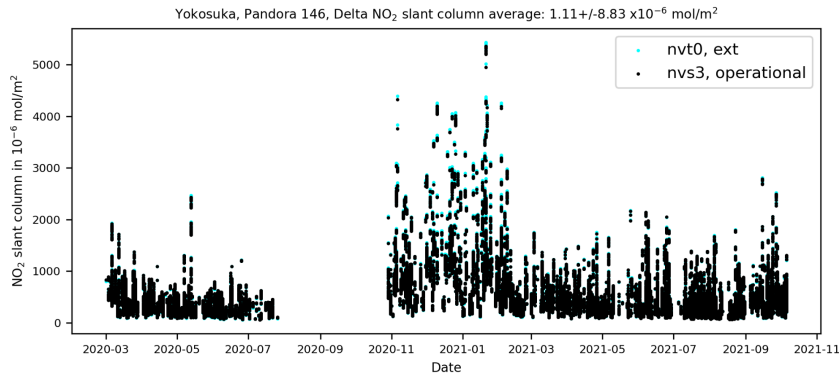
Wakkerstroom



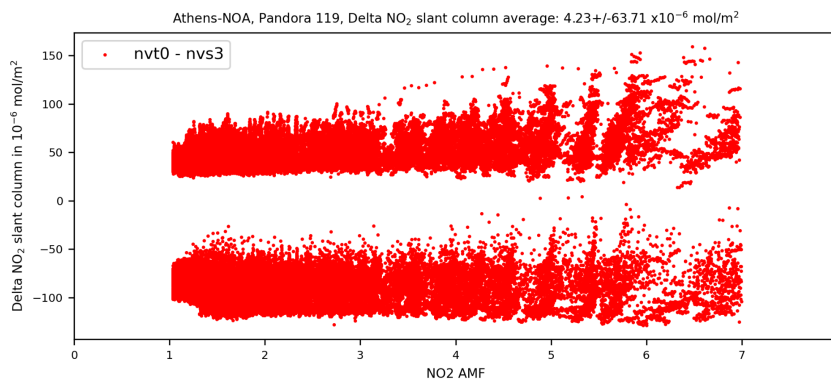
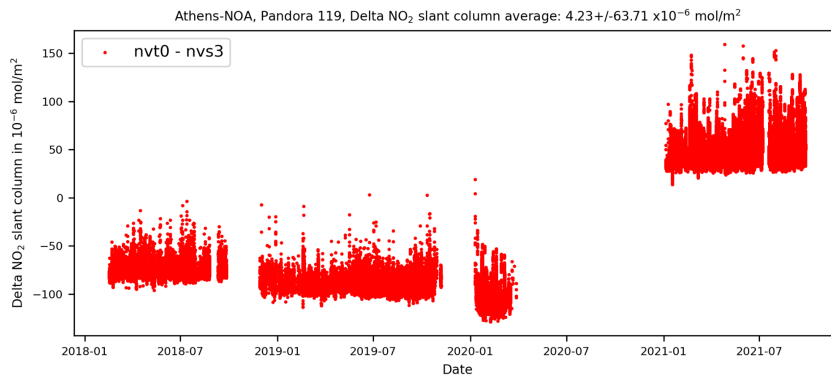
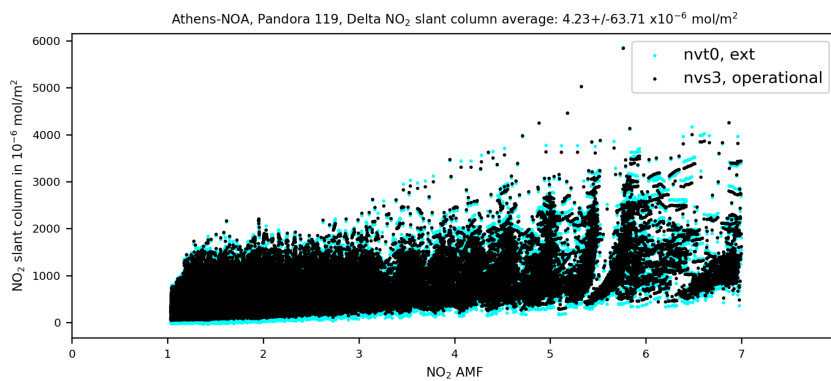
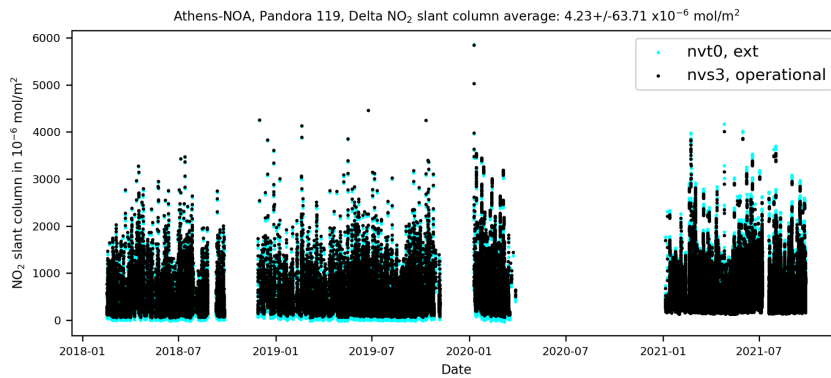
HoustonTX



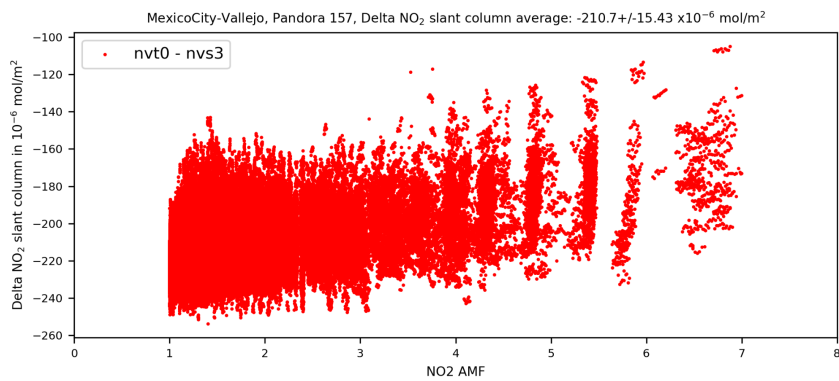
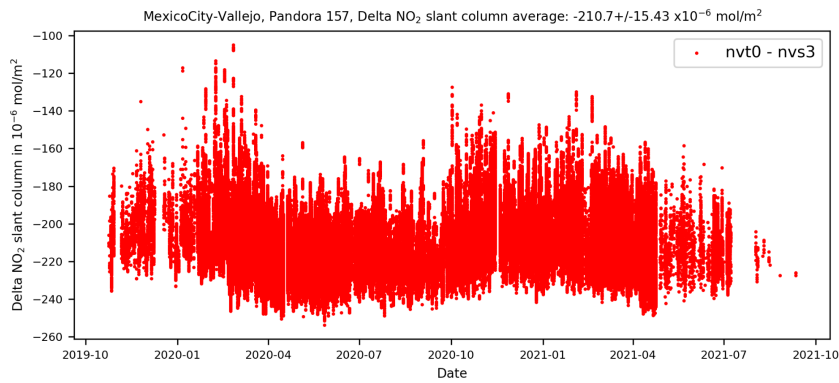
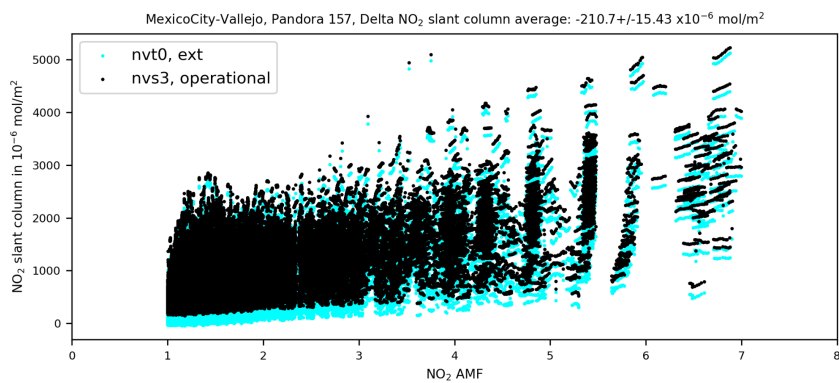
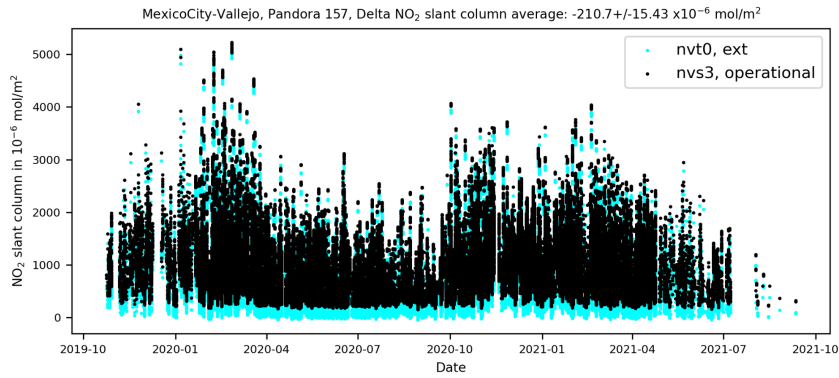
Yokosuka



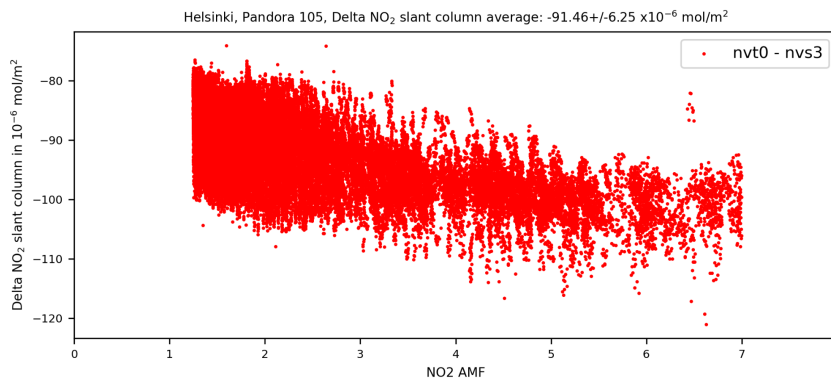
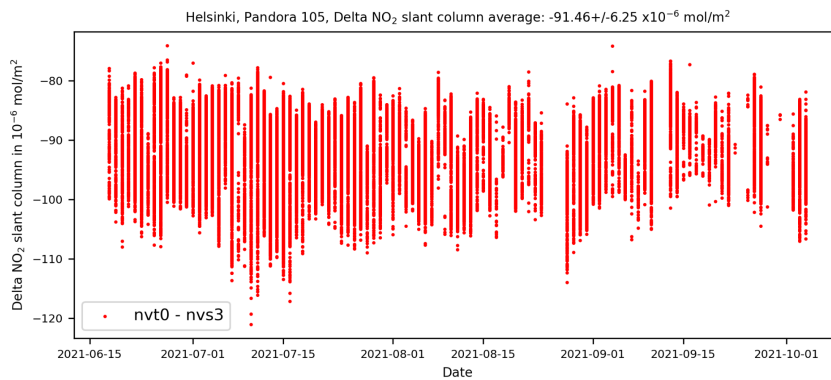
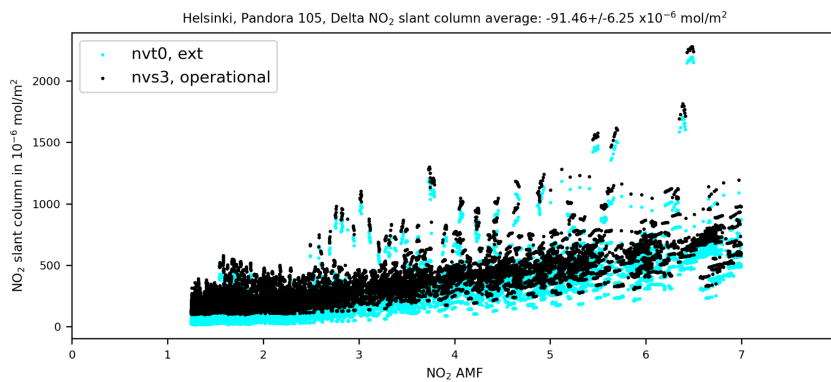
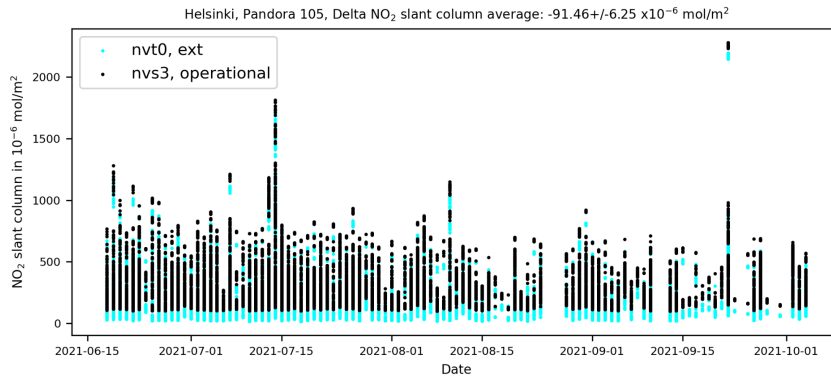
Athens



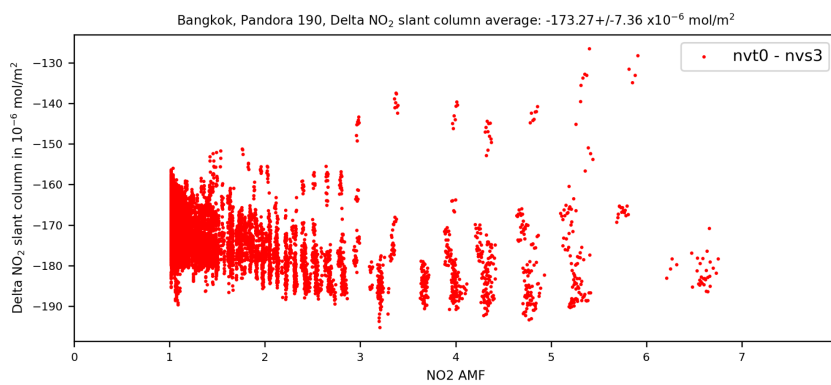
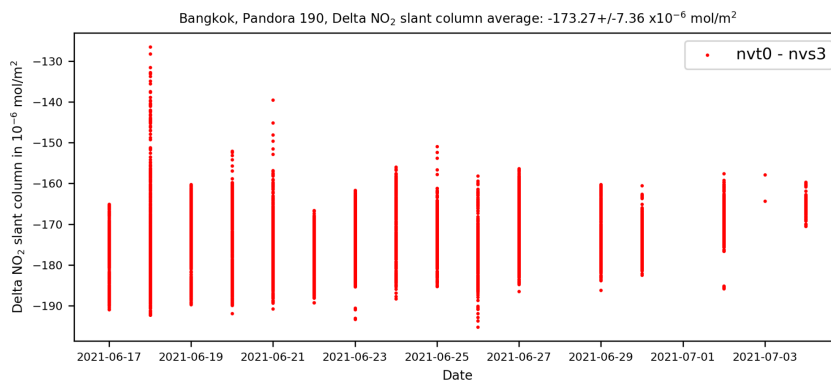
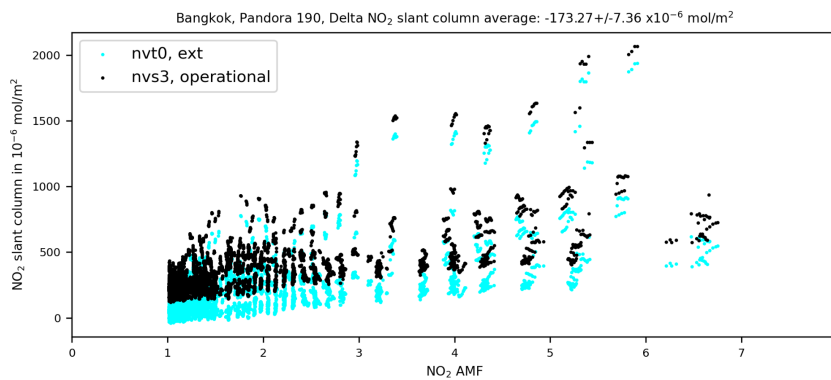
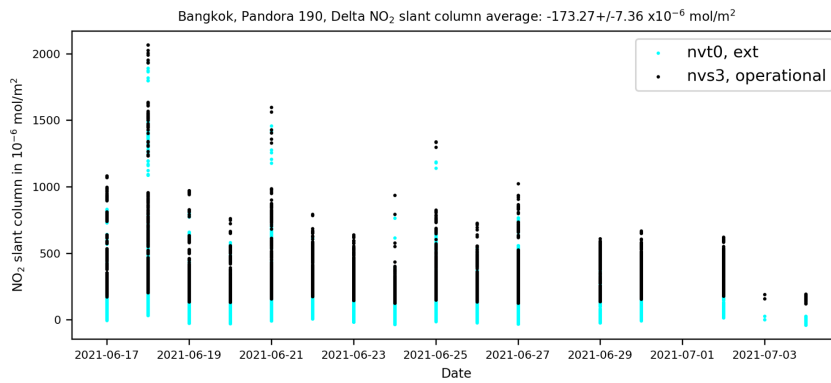
MexicoCity



Helsinki



Bangkok



Ulsan

