



# **Pandonia Global Network Data Products Readme Document**

Version 1.8-10, 20<sup>th</sup> Jan 2025

	<b>Name</b>	<b>Company</b>
<b>prepared by</b>	Alexander Cede	LuftBlick, NASA/GSFC
	Martin Tiefengraber	LuftBlick
	Manuel Gebetsberger	LuftBlick
	Elena Spinei Lind	Virginia Tech

# Contents

<b>Document Change Record</b>	<b>2</b>	<b>5</b>	<b>Uncertainty and data quality information</b>	<b>32</b>
<b>Acronyms and Abbreviations</b>	<b>4</b>		5.1 Uncertainty types . . . . .	32
<b>1 Introduction</b>	<b>5</b>		5.2 Total uncertainty for operational direct sun products . . . . .	33
1.1 Document overview . . . . .	5		5.3 Data Quality Flags . . . . .	35
1.2 Applicable Documents . . . . .	6		5.4 Not covered by uncertainty or DQF . . . . .	35
1.3 Reference Documents . . . . .	6		<b>6 Special notes</b>	<b>37</b>
<b>2 Readme pages for each retrieval setup</b>	<b>9</b>		6.1 Note on spatially heterogeneous trace gases . . . . .	37
2.1 PGN data taxonomy . . . . .	12		6.2 Note on retrievals in the short UV range . . . . .	38
2.2 fuh4 . . . . .	13		6.3 Note on weak absorbers . . . . .	38
2.3 fuh5 . . . . .	14		6.3.1 Weak absorbers in direct observation . . . . .	38
2.4 fus5 . . . . .	15		6.3.2 Weak absorbers in sky observation . . . . .	38
2.5 nvh2 . . . . .	16			
2.6 nvh3 . . . . .	17			
2.7 nvs0 . . . . .	18			
2.8 nvs1 . . . . .	19			
2.9 nvs2 . . . . .	20			
2.10 nvs3 . . . . .	21			
2.11 ous1 . . . . .	22			
2.12 out0 . . . . .	23			
2.13 out1 . . . . .	24			
2.14 out2 . . . . .	25			
2.15 sus1 . . . . .	26			
2.16 wvt1 . . . . .	27			
<b>3 Cross section sources</b>	<b>28</b>			
<b>4 Viewing geometries</b>	<b>29</b>			
4.1 Direct sun or moon . . . . .	29			
4.2 Profile . . . . .	30			
4.3 Satellite nadir view . . . . .	31			



## Document Change Record

Issue	Date	Section	Comments
1	31 <sup>st</sup> Dec 2019	All	First version
1.8-1	31 <sup>st</sup> Dec 2020	All	Rename as a readme document, re-structure it showing the r-codes in alphabetical order, modify all parts accordingly, add proper section on uncertainty and data quality
1.8-2	20 <sup>th</sup> Jan 2021	2	Add some representative figures to some readme pages; explained how to use GEOMS data
1.8-3	29 <sup>th</sup> Jan 2021	All	Update r-codes and cross sections table, change logo, expand uncertainty description
1.8-4	7 <sup>th</sup> Oct 2021	All	Added new r-codes: fuh5, fus5, nvh2, nvh3, nvs3, ous1, out2, sus1 due to updated quality limits. Added processor requirement to tables. Added info on data formats in introduction.
1.8-5	31 <sup>st</sup> Dec 2021	2 (page 9)	Resolution change fitting polynomial was included, in the description and in the fitting setup details.
1.8-6	31 <sup>st</sup> Dec 2022	2 (page 9), 2.11 (page 22)	Readme page for “nvh0” removed (processing temporarily discontinued), typo correction.
1.8-7	31 <sup>st</sup> May 2023	2.16 (page 27)	First official direct sun H <sub>2</sub> O r-code added (wvt1).
1.8-8	30 <sup>th</sup> Nov 2023	2.11 (page 22), 2.14 (page 25), 2.15 (page 26), 6.2 (page 38))	Additional information about the usage of rout2p1-8 and rsus1p1-8. Minor updates for ous1 and special notes for UV retrieved products.

Issue	Date	Section	Comments
1.8-9	30 <sup>th</sup> Nov 2024	2.11 (page 22), 2.14 (page 25), 2.15 (page 26), 6.2 (page 38), 5.3 (page 35))	Additional information about the usage of rout2p1-8 and rsus1p1-8. Minor updates for ous1 and special notes for UV retrieved products. Update on the description of the DQFs.
1.8-10	20 <sup>th</sup> Jan 2025	2.1 (page 12), 5.2 (page 33), 5.4 (page 35)	Section about naming guidelines and streamlining across the document (e.g. MAXDOAS changed to Profile). Uncertainty statistics for operational products have been added. Minor addition to aspects not covered by uncertainty.



## Acronyms and Abbreviations

AMF	Air Mass Factor
AOD	Aerosol Optical Depth
ATBD	Algorithm Theoretical Basis Document
BlickP	Blick Processing Software
BrO	Bromine oxide
DQ	Data Quality
DQF	Data Quality Flag
DU	Dobson Units
EffTemp	Effective temperature
EGL	Effective Ground Location
ESA	European Space Agency
FOV	Field of View
FRM4AQ-2	Fiducial Reference Measurements for Air Quality 2
FWHM	Full Width Half Maximum
GEOMS	Generic Earth Observation Metadata Standard
$h_{EFF}$	Effective Height
H <sub>2</sub> O	Water vapor
HCHO	Formaldehyde
HONO	Nitrous acid
L2	Level 2
MAXDOAS	Multi Axis Differential Optical Absorption Spectroscopy
NO <sub>2</sub>	Nitrogen dioxide
O <sub>2</sub> O <sub>2</sub>	Oxygen dimer
O <sub>3</sub>	Ozone
OIO	Iodine dioxide
PGN	Pandonia Global Network
ProcVers	BlickP processor version
QC	Quality Control
r-code	Retrieval Setup Code
rms	Root Mean Square
SI	International System of Units
SNR	Signal to Noise Ratio
SO <sub>2</sub>	Sulfur dioxide

SurfDens	Surface number density
TotCol	Total column amount
TropCol	Tropospheric column amount
TropProf	Tropospheric profile
U	Total uncertainty
$U_C$	Common uncertainty
$U_I$	Independent uncertainty
$U_S$	Structured uncertainty
USS	Unwanted Spectral Signal
UV	Ultraviolet
ZA	Zenith angle



# 1 Introduction

## 1.1 Document overview

This is the official "Readme document" for the Level 2 (L2) data products from the Pandora Global Network (PGN). It is also deliverable 8 (D8) of the ESA project "Fiducial Reference Measurements for Air Quality 2" (FRM4AQ-2) [1, 2] and gives guidelines on the correct use of the PGN data. Key questions answered include:

- What do the data represent?
- Which part of the atmosphere is sampled?
- How should the data be used?
- What shortcomings might the data have?

Each retrieval version is identified by the "retrieval setup code" (or "r-code") and the Blick processing software (BlickP) processor version (ProcVers), which were used for the data processing. r-code and ProcVers can be found in this way:

- Native PGN data format: r-code and ProcVers appear in the L2 file name `PandoraXsY_LLL_L2_rRRRRpP.txt`. X is the instrument number, Y the spectrometer number, LLL the location name, RRRR the r-code and P the ProcVers.
- GEOMS data format: r-code and ProcVers appear in the GEOMS file attribute "DATA\_PROCESSOR" `rRRRRpP`. RRRR is the r-code and P the ProcVers.

The r-code defines the detailed settings for the spectral fitting (e.g. the length of the fitting window) and the algorithm-type used for the retrieval. Different r-codes in general cause the values of L2 data to change.

When changing the ProcVers, there are two effects on the L2 data:

### Data format changes

The data format of each ProcVers can be found in the respective Blick Software Suite Manual on the PGN-webpage under /Downloads/Manuals. Within the same ProcVers the format does not change, while newer ProcVers typically have more

output columns than previous versions. E.g., from ProcVers 1.7 to 1.8 more columns describing the data uncertainty have been added.

### Retrieval algorithm may change

It is possible that retrieval algorithm of a newer ProcVers is modified compared to the previous one. E.g., a newer ProcVers might use a different temperature climatology or different formula for the Rayleigh correction in the sky algorithms. Therefore the values of data from different ProcVers might vary even if the same r-code was used.

A PGN data product can have one of the following "Product status":

- Product status "official": Data are currently an official product of the PGN.
- Product status "disused": Data are a former official product of the PGN, but are not an official product anymore.
- Product status "unvalidated": Data are not yet officially validated.
- Product status "develop": Data are in research stage and only produced for testing purposes.
- Product status "unusable": Data should not be used, since they are typically only for an additional output gas in the spectral fitting, for which the wavelength range is not optimized.

Section 2 gives the so-called "Readme pages" for those r-codes with product status "official" or "disused" or also "unvalidated" in case it is expected that this r-code will soon become an official one. The r-codes are sorted in alphabetical order (numbers are after letters, i.e. "nvs0" would be after "nvsA"). Hence the user of a certain PGN data product can just look up the r-code of the data file, search for it in the content section of this document, and then jump directly to the respective readme page.

The sections after section 2 give additional information with respect to the understanding of the data and are referenced in the readme pages.

Since this document is intended to give a concise overview for each data product, the user is referred to the Algorithm Theoretical Basis Document (ATBD) for more details. The ATBD is given as section "Algorithms Theoretical Basis" in the Blick Software Suite Manual [3].



## 1.2 Applicable Documents

- [1] Fiducial Reference Measurements for Air Quality [Proposal], LuftBlick Proposal 201805DEV, Issue 1, 2018.
- [2] Fiducial Reference Measurements for Air Quality [Contract and Statement of Work], ESA Contract No. 4000125841/18/I-NS, 2018.
- [3] A. Cede. *Manual for Blick Software Suite 1.8, Version 1-8-1*, 2020. URL [https://www.pandonia-global-network.org/wp-content/uploads/2020/12/BlickSoftwareSuite\\_Manual\\_v1-8-1.pdf](https://www.pandonia-global-network.org/wp-content/uploads/2020/12/BlickSoftwareSuite_Manual_v1-8-1.pdf).
- [4] M. Gebetsberger, A. Cede, and M. Tiefengraber. Fiducial Reference Measurements for Air Quality, LuftBlick Report 2019008: TN on Data Quality Flaggng Generic Procedure Evolution, 2022.
- [15] M. Müller, M. Tiefengraber, and A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2016011: Validation reports, 2016.
- [25] M. Tiefengraber and A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2016001: Report on Feasibility to Retrieve Trace Gases other than O3 and NO2 with Pandora, 2016.

## 1.3 Reference Documents

- [9] B. A. Bodhaine, N. B. Wood, E. G. Dutton, and J. R. Slusser. On Rayleigh Optical Depth Calculations. *Journal of Atmospheric and Oceanic Technology*, 16:1854, November 1999. doi: 10.1175/1520-0426(1999)016<1854:ORODC>2.0.CO;2.
- [10] Alexander Cede, Eduardo Luccini, Liliana Nuñez, Rubén D Piacentini, Mario Blumthaler, and Jay R Herman. Toms-derived erythemal irradiance versus measurements at the stations of the argentine uv monitoring network. *Journal of Geophysical Research: Atmospheres*, 109(D8), 2004.
- [11] J. F. Grainger and J. Ring. Anomalous Fraunhofer line profiles. *Nature*, 193: 762, 1962.
- [12] J. Herman, A. Cede, E. Spinei, G. Mount, M. Tzortziou, and N. Abuhassan. NO<sub>2</sub> column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation. *Journal of Geophysical Research (Atmospheres)*, 114: D13307, July 2009. doi: 10.1029/2009JD011848.
- [13] Laura M Judd, Jassim A Al-Saadi, Scott J Janz, Matthew G Kowalewski, R Bradley Pierce, James J Szykman, Lukas C Valin, Robert Swap, Alexander Cede, Moritz Mueller, et al. Evaluating the impact of spatial resolution on tropospheric no<sub>2</sub> column comparisons within urban areas using high-resolution airborne data. *Atmospheric Measurement Techniques*, 12(11):6091–6111, 2019.
- [14] Richard Meller and Geert K Moortgat. Temperature dependence of the absorption cross sections of formaldehyde between 223 and 323 k in the wavelength range 225–375 nm. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 105(D6):7089–7101, 2000.
- [15] M. Müller, M. Tiefengraber, and A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2016011: Validation reports, 2016.
- [16] David A Newnham and John Ballard. Visible absorption cross sections and integrated absorption intensities of molecular oxygen (o<sub>2</sub> and o<sub>4</sub>). *Journal of*



- Geophysical Research: Atmospheres* (1984–2012), 103(D22):28801–28815, 1998.
- [17] James C Owens. Optical refractive index of air: dependence on pressure, temperature and composition. *Applied Optics*, 6(1):51–59, 1967.
- [18] Laurence S Rothman, Iouli E Gordon, Yury Babikov, Alain Barbe, D Chris Benner, Peter F Bernath, Manfred Birk, Luca Bizzocchi, Vincent Boudon, Linda R Brown, et al. The hitran2012 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 130:4–50, 2013.
- [19] A Saiz-Lopez, RW Saunders, DM Joseph, SH Ashworth, and JMC Plane. Absolute absorption cross-section and photolysis rate of  $\text{I}_2$ . *Atmospheric Chemistry and Physics*, 4(5):1443–1450, 2004.
- [20] A. Serdyuchenko, V. Gorshchev, M. Weber, W. Chehade, and J. P. Burrows. High spectral resolution ozone absorption cross-sections - Part 2: Temperature dependence. *Atmospheric Measurement Techniques Discussions*, 6:6613–6643, July 2013. doi: 10.5194/amtd-6-6613-2013.
- [21] Peter Spietz. *Absorption Cross Sections for Iodine Species of Relevance to the Photolysis of Mixtures of  $\text{I}_2$  and  $\text{O}_3$  and for the Atmosphere*. PhD thesis, University of Bremen, 2005.
- [22] E. Spinei, M. Tiefengraber, M. Mueller, M. Gebetsberger, A. Cede, L. Valin, J. Szykman, A. Whitehill, F. Kostakis, A. Santos, N. Abbuhasan, X. Zhao, V. Fioletov, S. Chi Lee, and R. Swap. Effect of Polyoxymethylene (POM-H Delrin) offgassing within Pandora head sensor on direct sun and multi-axis formaldehyde column measurements in 2016 - 2019. *Atmospheric Measurement Techniques Discussions*, in review, 2020.
- [23] J Stutz, ES Kim, U Platt, P Bruno, C Perrino, and A Febo. Uv-visible absorption cross sections of nitrous acid. *Journal of Geophysical Research: Atmospheres* (1984–2012), 105(D11):14585–14592, 2000.
- [24] Ryan Thalman and Rainer Volkamer. Temperature dependent absorption cross-sections of  $\text{O}_2\text{-O}_2$  collision pairs between 340 and 630 nm and at atmospherically relevant pressure. *Physical chemistry chemical physics*, 15(37):15371–15381, 2013.
- [25] M. Tiefengraber and A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2016001: Report on Feasibility to Retrieve Trace Gases other than  $\text{O}_3$  and  $\text{NO}_2$  with Pandora, 2016.
- [26] M. Tiefengraber and A. Cede. More accurate Pandora total ozone columns by improved laboratory calibration and simultaneous retrieval of effective ozone temperature. Quadrennial Ozone Symposium 2016, Edinburgh, United Kingdom, 2016.
- [27] Jan-Lukas Tirpitz, Udo Frieß, François Hendrick, Carlos Alberti, Marc Allaart, Arnoud Apituley, Alkis Bais, Steffen Beirle, Stijn Berkhout, Kristof Bognar, Tim Bösch, Ilya Bruchkouski, Alexander Cede, Ka Lok Chan, Mirjam den Hoed, Sebastian Donner, Theano Drosoglou, Caroline Fayt, Martina M. Friedrich, Arnoud Frumau, Lou Gast, Clio Gielen, Laura Gomez-Martín, Nan Hao, Arjan Hensen, Bas Henzing, Christian Hermans, Junli Jin, Karin Kreher, Jonas Kuhn, Johannes Lampel, Ang Li, Cheng Liu, Haoran Liu, Jianzhong Ma, Alexis Merlaud, Enno Peters, Gaia Pinardi, Ankie Piters, Ulrich Platt, Olga Puentedura, Andreas Richter, Stefan Schmitt, Elena Spinei, Deborah Stein Zweers, Kimberly Strong, Daan Swart, Frederik Tack, Martin Tiefengraber, René van der Hoff, Michel van Roozendaal, Tim Vlemmix, Jan Vonk, Thomas Wagner, Yang Wang, Zhuoru Wang, Mark Wenig, Matthias Wiegner, Folkard Wittrock, Pinhua Xie, Chengzhi Xing, Jin Xu, Margarita Yela, Chengxin Zhang, and Xiaoyi Zhao. Intercomparison of MAX-DOAS vertical profile retrieval algorithms: studies on field data from the CINDI-2 campaign. *Atmospheric Measurement Techniques*, 14(1):1–35, January 2021. doi: 10.5194/amt-14-1-2021. URL <https://doi.org/10.5194/amt-14-1-2021>.
- [28] A. C. Vandaele, C. Hermans, P. C. Simon, M. Carleer, R. Colin, S. Fally, M. F. Mérienne, A. Jenouvrier, and B. Coquart. Measurements of the  $\text{NO}_2$  absorption cross-section from  $42,000\text{ cm}^{-1}$  to  $10,000\text{ cm}^{-1}$  (238–1000 nm) at 220 K and 294 K. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 59:171–184, May 1998. doi: 10.1016/S0022-4073(97)00168-4.
- [29] Ann Carine Vandaele, Paul C Simon, Jean Michel Guilmot, Michel Carleer, and Réginald Colin.  $\text{SO}_2$  absorption cross section measurement in the uv us-



ing a fourier transform spectrometer. *Journal of Geophysical Research: Atmospheres* (1984–2012), 99(D12):25599–25605, 1994.

- [30] Ann Carine Vandaele, Christian Hermans, and Sophie Fally. Fourier transform measurements of so<sub>2</sub> absorption cross sections: II.: Temperature dependence in the 29 000–44 000 cm<sup>-1</sup> (227–345 nm) region. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 110(18):2115–2126, 2009.
- [31] David M Wilmouth, Thomas F Hanisco, Neil M Donahue, and James G Anderson. Fourier transform ultraviolet spectroscopy of the a  $2\pi 3/2 \rightarrow 2\pi 3/2$  transition of bro. *The Journal of Physical Chemistry A*, 103(45):8935–8945, 1999.
- [32] X. Zhao, V. Fioletov, A. Cede, J. Davies, and K. Strong. Accuracy, precision, and temperature dependence of Pandora total ozone measurements estimated from a comparison with the Brewer triad in Toronto. *Atmospheric Measurement Techniques*, 9(12):5747–5761, 2016. doi: 10.5194/amt-9-5747-2016. URL <https://www.atmos-meas-tech.net/9/5747/2016/>.





## 2 Readme pages for each retrieval setup

Each of the following subsections presents the readme page for all retrieval setups producing the data products. It starts with a description of the related data product and is followed by a table with the major retrieval settings such as the cross sections used, the wavelength range for the spectral fitting etc. The meaning of the entries in the tables is explained here:

### Code creator(s)

The name of the person having created the r-code plus the creation date. If different persons have created the respective f-codes (see ATBD, table "f-codes") used by the r-code, then their names are added as well.

### QC limits creator(s)

The name of the person having created the qr-code (see ATBD, table "qr-codes") plus the creation date. The qr-code sets the limits to determine the data quality (DQ). If different persons have created the respective qf-codes (see ATBD table, "qf-codes") used by the f-code, then their names are added as well.

### Data product(s)

A comma-separated string for each data product. Each substring gives the acronym of the species followed by the product type and the unit in square brackets. The product type can be:

- TotCol: total vertical column amount of the trace gas between the surface and the top of the atmosphere. Starting with ProcVers 1.8, for most products the quantity is given in the SI-unit for column density, which is mole per square meter [ $\text{mol}/\text{m}^2$ ] and each of the uncertainty types listed in section 5.1 is given.
- TropCol: tropospheric vertical column amount of the trace gas between the surface and approximately 4 km above the surface in the same unit as TotCol. For TropCols only the independent uncertainty is given at present (see section 5.1).
- SurfDens: the average number density of the trace gas in approximately the first 50 m above the surface. Starting with ProcVers 1.8, this quantity is given

in the SI-unit for concentration, which is mole per cubic meter [ $\text{mol}/\text{m}^3$ ]. For SurfDens only the independent uncertainty is given at present (see section 5.1).

- TropProf: partial column amounts of the trace gas for several layers between the surface and approximately 4 km above the surface in the same unit as TotCol. For TropProfs no output uncertainty is given at present. This output does not exist before ProcVers 1.8.
- EffTemp: Effective temperature of the trace gas column in Kelvin [K]. This is the average of the trace gas temperature along its slant path weighted with its concentration along the slant path (more info in the ATBD, section "Table 'Trace gases'"). Starting with ProcVers 1.8, each of the uncertainty types listed in section 5.1 is given for EffTemp.

E.g. Data products "O3 TotCol [ $\text{mol}/\text{m}^2$ ], O3 EffTemp [K]" means that this r-code has outputs 2 data products, the total ozone column and the effective ozone temperature.

### Processor requirement

This gives the ProcVers needed to calculate the respective r-code. When such r-code is run on a ProcVers not listed, then BlickP typically does not produce any data. Note that when an r-code is run on an older previous ProcVers, it is possible that not all output products as listed in entry "Output product(s)" might be given. E.g., r-codes run with ProcVers 1.7 will not give any TropProf.

### Product status

A product status as described in section 1 for each output product. A former product status may be added in parenthesis.

### Observation mode

This describes where the instrument was pointing during the measurements. The observation mode can be "Direct sun" or "Direct moon" (see section 4.1) or "Profile" (see section 4.2).

Measurements in a direct observation mode are processed with the "L2 Direct Algorithm" with output products TotCol and eventually also EffTemp (see ATBD,



section "L2 Direct Algorithm"). Measurements in a profile mode are processed with the "L2 Air-Ratio Sky Algorithm" with output products TropCol, SurfConc and TropProf (see ATBD, section "L2 Air-Ratio Sky Algorithm").

### Filters used

Pandora uses a filterwheel, where different spectral filter can be placed, specifically a U340 and a BP300 filter. U340 is a colored glass bandpass filter with peak transmission at 340 nm. BP300 is an interference bandpass filter with peak transmission at 300 nm. If none of these filters is in place, the measurement is called OPEN. This setting indicates, on which type of filter settings the respective r-code is applied.

### Effective height

This (comma-separated) string is only used for observation modes "Direct sun" or "Direct moon" and gives the assumed effective heights for each output trace gas used to calculate the direct air mass factor (AMF, see section 4.1). For observation mode "profile" it is not given.

Each sub-string is either a value in km or a string representing a profile type from a climatology. The options for the climatology are listed in section "Climatologies" of the ATBD. The string can be followed by an uncertainty estimation in round parenthesis. This estimation can be a value in km or a string representing a profile type. E.g. "O3(2.4)" would mean the effective height is taken from the ozone climatology and its uncertainty is 2.4 km. If the uncertainty estimation is not given and a string representing a profile type is set for the effective height, then the uncertainty estimation is taken from the same climatology too, i.e. "O3" has the same effect as "O3(O3)". If the uncertainty estimation is not given and a value in km is set for the effective height, the retrieval uses half of the effective height as uncertainty, i.e. "7.2" has the same effect as "7.2(3.6)". Sub-strings can also be combined using "+" for the direct sun or moon observation mode, e.g. "BL+STRAT". In this case different effective height-estimations are used to get the best estimation of the AMF.

### Number of fitting windows

This is either 1 or 2. In the latter case 2 separate fitting processes are performed and the final output data are a combination of the results of both of them (see ATBD, section "L2 Direct Algorithm"). In this case the remaining entries of the table can be strings separated by a semicolon, but only if the 2 fitting processes use different

settings. The 2nd part of the string, after the semicolon, is then written in italic for easier distinction.

### Reference(s)

The reference spectra used in the spectral fitting. It is one of the following options (more detail in the ATBD, table "Options for 'Reference'"):

- Theoretical reference spectrum: an extraterrestrial spectrum from another source (i.e. not measured by the Pandora unit), convoluted with the Pandora filter function and sensitivity. BlickP uses for this purpose a high resolution extraterrestrial spectrum from 270 nm to 1000 nm merged from different sources in a similar way as described in *Bernhard et al.* [7]. For PGN data products using this theoretical reference, there is no common uncertainty output given (see section 5.1). This does NOT mean, that such common uncertainty does not exist, it is just not quantified in the current version of BlickP yet. A rough estimation of the common uncertainty for these products is given in the specific readme pages.
- Synthetic reference spectrum: a spectrum, which is usually the average over several spectra measured by the Pandora unit and corrected for the estimated total optical depth from the different atmospheric extinction processes included in it. The uncertainty is these estimated optical depths, which are determined during the instrument calibration, is the driver for the common uncertainty of the product. Typically the calibration works better at stations, which are not always polluted, and consequently works worse at continuously polluted places such as large cities, which is then reflected in the common uncertainty output.
- Largest pointing zenith angle: the spectrum measured by the instrument at the highest pointing zenith angle of the measurement sequence is used as a reference. This measurement is typically below 3° above the horizon. Since BlickP does not assume any systematic instrumental changes within the measurements of one sequence, the obtained column amount with this reference do not have a common uncertainty output.

### Wavelength window(s)

The lower and upper limits of the wavelength window(s) used in the spectral fitting



in nanometer [nm].

### Order(s) of polynomials

The orders of the smoothing polynomial, "SMO", offset polynomial, "OFFS", wavelength change polynomial, "WLC", and resolution change polynomial "RSC" used in the spectral fitting (see ATBD, section "L2Fit Algorithm - Spectral Fitting"). E.g. "SMO 4, OFFS 0, WLC 1, RSC 0" means a 4th order smoothing, 0 order offset, 1st order wavelength change polynomial and a zero order resolution correction polynomial have been used. A value of "-1" means that the respective polynomial is not used in the fitting.

### Fitted gases

A comma separated string listing the trace gases used in the spectral fitting, the source of the cross sections and the treatment of the temperature. Each substring has format "GGG (TTT)". GGG refers to an entry of table 2, which lists the source of the cross sections for the respective trace gas. TTT describes the temperature treatment and can be:

- "T-fit": the effective temperature is fitted.
- A temperature value: this means the cross sections at this fixed temperature (e.g. 225.0 K ) are used in the fitting.
- "XX-clim": the spectral fitting uses the temperature from climatology "XX" (see ATBD, section "Climatologies"). E.g. "O3-clim" means the effective ozone temperature climatology is used.

### Ring

This is "Fitted", if the Rotational Raman Scattering [11] was included in the spectral fitting, and "Not fitted" otherwise.

### Molecular scattering

This is "Subtracted", if the molecular scattering cross sections [9, 17] have been subtracted before the spectral fitting, and "Not subtracted" otherwise.

### Uncertainty

This entry indicates how the uncertainty was included in the spectral fitting. There

are 3 options (for more detail in ATBD, section "L2Fit Algorithm - Spectral Fitting"):

- NO: no uncertainty was included in the spectral fitting.
- INSTR: the independent instrumental uncertainty was included in the spectral fitting.
- MEAS: uncertainty based on the standard deviation over the measurement cycles was included in the spectral fitting.

### AMF limits

AMF limits and the other entries in the tables including the word "limits" are threshold values used to determine the data quality. They have two comma-separated values, which represent the limits for this parameter to change the last digit of the Data Quality Flag (DQF) from 0 to 1 and from 1 to 2 respectively. More information on the data quality is given in the ATBD, section "Data quality flags" and in section 5.3 here. AMF limits refers to the maximum allowed direct AMF for a given DQ. E.g. "AMF limits: 7, 14" means that measurement taken at AMF>7 receive a unit-value of the DQF of 1 and those at AMF>14 even a DQF unit-value of 2.

### AtmVar limits

AtmVar stands for the atmospheric variability. This is a percentage <100% indicating the magnitude of the atmospheric variability during the measurements. Small numbers mean the atmosphere was very stable. 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. More details can be found in the ATBD, section "L1 Uncertainty".

### Wavelength shift limits

This refers to the maximum allowed wavelength shift obtained in the spectral direct for a given DQ in nanometer [nm].

### wrms limits

This refers to the maximum allowed values for the normalized rms of the fitting residuals weighted with the independent uncertainty (see also ATBD, section "Residuals").



## 2.1 PGN data taxonomy

**Table 1:** Overview of the hierarchical data naming taxonomy used for PGN data, from fundamental elements (species, product types) to higher-level groupings (data sets, collections). Each row lists the term, an illustrative example, and a brief explanation of its meaning.

Term	Example	Meaning
<b>Species</b>	NO <sub>2</sub>	Short name (often the chemical formula) of the atmospheric species considered.
<b>Data product type</b>	TotCol	The type of the retrieved value (e.g., total column (TotCol), surface number density (SurfDens)).
<b>Data product</b>	NO <sub>2</sub> TotCol	Combination of the <i>Species</i> and the <i>Data product type</i> . Note that data products can have multiple versions.
<b>Retrieval setup</b>	rnvs3	Summarizes the settings used for the retrieval (r-code), following the nomenclature in <i>Cede</i> [3].
<b>Processor version</b>	p1-8	Blick processing software version ("processor").
<b>Retrieval version</b>	rnvs3p1-8	A specific combination of <i>Retrieval setup</i> and <i>Processor version</i> .
<b>Versioned data product</b>	NO <sub>2</sub> TotCol rnvs3p1-8	A <i>Data product</i> referencing a specific <i>Retrieval version</i> .

Term	Meaning
<b>Data set</b>	A time series of a <i>Versioned data product</i> for a specific deployment (location + period).
<b>Data set group</b>	A group of <i>Data sets</i> for a single deployment (e.g., combining multiple species, time ranges, or measurement types).
<b>Data collection</b>	A larger collection of <i>Data set groups</i> (e.g., from multiple deployments or sites) gathered for a campaign or project.



## 2.2 fuh4

fuh4 provides HCHO TropCols, SurfDens and TropProfs. A comparison of the data from this algorithm with other profile retrievals can be found in *Tirpitz et al.* [27].

On 21 May 2021 it has been replaced by nvs1 and is now "disused".

r-code	fuh4
Code creator	Elena Spinei Lind, 30 Dec 2020
DQ limits creator	Martin Tiefengraber, 11 May 2017
Data products	HCHO TropCol [mol/m <sup>2</sup> ], HCHO SurfConc [mol/m <sup>3</sup> ], HCHO TropProf [mol/m <sup>2</sup> ]
Processor requirement	1.7 and higher
Product status	disused, disused, disused
Observation mode	Profile
Filters used	OPEN, U340
Number of fitting windows	1
Reference	Largest pointing zenith angle
Wavelength window	328.5 nm - 359.0 nm
Order of polynomials	SMO 4, OFFS 1, WLC 1, RSC -1
Fitted gases	O3-2 (BL-clim), NO2-1 (BL-clim), HCHO-1 (BL-clim), O2O2-1 (BL-clim)
Ring	Fitted
Molecular scattering	Not subtracted
Uncertainty	INSTR
AMF limits	6, 10
AtmVar limits	20, 40
Wavelength shift limits	0.05, 0.1
wrms limits	1e-3, 3e-3



## 2.3 fuh5

fuh5 is as fuh4, but uses of updated quality flag thresholds based on *Gebetsberger et al.* [4].

fuh5 is the first official PGN code for HCHO TropCols, SurfDens and TropProfs.

r-code	fuh5
Code creator	Elena Spinei Lind, 30 Dec 2020
DQ limits creator	Manuel Gebetsberger, 21 May 2021
Data products	HCHO TropCol [mol/m <sup>2</sup> ], HCHO SurfConc [mol/m <sup>3</sup> ], HCHO TropProf [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	official, official, official
Observation mode	Profile
Filters used	OPEN, U340
Number of fitting windows	1
Reference	Largest pointing zenith angle
Wavelength window	328.5 nm - 359.0 nm
Order of polynomials	SMO 4, OFFS 1, WLC 1, RSC -1
Fitted gases	O3-2 (BL-clim), NO2-1 (BL-clim), HCHO-1 (BL-clim), O2O2-1 (BL-clim)
Ring	Fitted
Molecular scattering	Not subtracted
Uncertainty	INSTR
AMF limits	6, 10
AtmVar limits	81, 85
Wavelength shift limits	0.05, 0.1
wrms limits	1.12e-3, 1.50e-3



## 2.4 fus5

fus5 provides HCHO TotCols. This code is only applied to instruments, which have no Delrin parts in the head sensor left (see *Spinei et al.* [22]). It is the first official PGN code for HCHO TotCols.

HCHO is a weak absorber and therefore the user should read section 6.3.1 before using fus5 data.

r-code	fus5
Code creator	Martin Tiefengraber, 30 Oct 2020
DQ limits creator	Manuel Gebetsberger, 21 May 2021
Output product	HCHO TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	official
Observation mode	Direct sun
Filters used	U340
Effective height	BL
Number of fitting windows	1
Reference	Synthetic reference spectrum
Wavelength window	322.5 nm - 359.2 nm
Order of polynomials	SMO 4, OFFS 0, WLC 0, RSC 0
Fitted gases	O3-2 (O3-clim), NO2-1 (BL-clim), SO2-2 (BL-clim), HCHO-1 (BL-clim), HONO-1 (BL-clim), BrO-1 (O3-clim)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	INSTR
AMF limits	6, 10
AtmVar limits	61, 70
Wavelength shift limits	0.05, 0.1
wrms limits	4.9e-4, 7.0e-4



## 2.5 nvh2

nvh2 provides NO<sub>2</sub> TropCols, SurfDens and TropProfs. A comparison of the data from this algorithm with other profile retrievals can be found in *Tirpitz et al.* [27].

On 21 May 2021 it has been replaced by nvs1 and is now "disused".

<b>r-code</b>	<b>nvh2</b>
Code creator	Elena Spinei Lind, 20 Jan 2017
DQ limits creator	Alexander Cede, 20 Jan 2017
Data products	NO <sub>2</sub> TropCol [mol/m <sup>2</sup> ], NO <sub>2</sub> SurfConc [mol/m <sup>3</sup> ], NO <sub>2</sub> TropProf [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	disused, disused, disused
Observation mode	Profile
Filters used	OPEN, U340
Number of fitting windows	1
Reference	Largest pointing zenith angle
Wavelength window	435.0 nm - 490.0 nm
Order of polynomials	SMO 4, OFFS 1, WLC 1, RSC -1
Fitted gases	O <sub>3</sub> -2 (BL-clim), NO <sub>2</sub> -1 (BL-clim), H <sub>2</sub> O-1 (BL-clim), O <sub>2</sub> O <sub>2</sub> -1 (BL-clim)
Ring	Fitted
Molecular scattering	Not subtracted
Uncertainty	INSTR
AMF limits	7, 14
AtmVar limits	81, 85
Wavelength shift limits	0.05, 0.1
wrms limits	6.9e-4, 9.1e-4





## 2.6 nvh3

nvh3 is as nvh2, but uses updated quality flag thresholds based on *Gebetsberger et al.* [4].

nvh3 is the first official PGN code for NO<sub>2</sub> TropCols, SurfDens and TropProfs.

<b>r-code</b>	<b>nvh3</b>
Code creator	Elena Spinei Lind, 20 Jan 2017
DQ limits creator	Manuel Gebetsberger, 21 May 2021
Data products	NO2 TropCol [mol/m <sup>2</sup> ], NO2 SurfConc [mol/m <sup>3</sup> ], NO2 TropProf [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	official, offical, official
Observation mode	Profile
Filters used	OPEN, U340
Number of fitting windows	1
Reference	Largest pointing zenith angle
Wavelength window	435.0 nm - 490.0 nm
Order of polynomials	SMO 4, OFFS 1, WLC 1, RSC -1
Fitted gases	O3-2 (BL-clim), NO2-1 (BL-clim), H2O-1 (BL-clim), O2O2-1 (BL-clim)
Ring	Fitted
Molecular scattering	Not subtracted
Uncertainty	INSTR
AMF limits	7, 14
AtmVar limits	81, 85
Wavelength shift limits	0.05, 0.1
wrms limits	6.9e-4, 9.1e-4



## 2.7 nvs0

nvs0 was the first code used for the official PGN total column NO<sub>2</sub> product. On 4 Nov 2019 it has been replaced by nvs1 and is now "disused".

All of the existing data with this code were processed with BlickP version 1.7, which does not have such a comprehensive uncertainty output as version 1.8. 1.7 still reports the NO<sub>2</sub> TotCol in DU, where 1 DU corresponds to 4.4615e-4 mol/m<sup>2</sup>. The total uncertainty for nvs0 data with DQF 0, which were processed with BlickP version 1.7, is typically estimated to 0.05 DU = 2.25e-5 mol/m<sup>2</sup> divided by the direct AMF, based on *Herman et al.* [12].

The main shortcomings of nvs0 data are:

- The effective height and temperature of the NO<sub>2</sub> profile are fixed to 7.2 km and 254.4 K respectively. These values assume an NO<sub>2</sub> profile with about 40% of the column in the stratosphere and 60% in the troposphere. However in real situations the tropospheric fraction can vary from 0% to >90% which would alter these estimations significantly. This creates biases in stations, which systematically have a different tropospheric fractions and could best be seen on rural or high altitude sites without pollution.
- The wrms limits are set in a to little rigorous way and therefore data got DQF 0 or 1, when they should have rather gotten higher DQFs.

NO<sub>2</sub> is a spatially heterogeneous trace gases and therefore the user should read section 6.1 before using nvs0 data.

r-code	nvs0
Code creator	Alexander Cede, 20 Jan 2017
DQ limits creator	Alexander Cede, 20 Jan 2017
Output product	NO2 TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.7 and higher
Product status	disused (official until 4 Nov 2019)
Observation mode	Direct sun
Filters used	OPEN
Effective height	7.2 km
Number of fitting windows	1
Reference	Synthetic reference spectrum
Wavelength window	400.0 nm - 440.0 nm
Order of polynomials	SMO 4, OFFS 0, WLC 1, RSC -1
Fitted gases	O3-1 (225.0 K), NO2-1 (254.5 K)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	MEAS
AMF limits	7, 14
AtmVar limits	20, 40
Wavelength shift limits	0.2, 0.5
wrms limits	2e-3, 5e-3

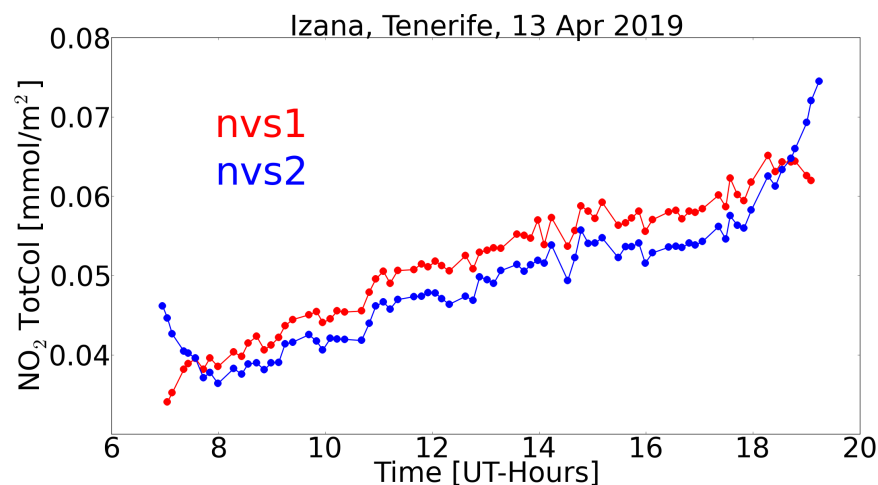


## 2.8 nvs1

nvs1 has replaced nvs0 as official PGN total column NO<sub>2</sub> product code on 4 Nov 2019, and was itself replaced by nvs2 on 20 Nov 2020. It differs from nvs0 in the thresholds for the wavelength shift and wrms limits to avoid that medium and low quality data "slip" into the segments of high and medium quality data respectively. All other aspects for the nvs1 data are the same as for nvs0 data.

The main shortcoming of nvs1 data is:

- The effective height and temperature of the NO<sub>2</sub> profile are fixed to 7.2 km and 254.4 K respectively. These values assume an NO<sub>2</sub> profile with about 40% of the column in the stratosphere and 60% in the troposphere. However in real situations the tropospheric fraction can vary from 0% to >90% which would alter these estimations significantly. This creates biases in stations, which systematically have a different tropospheric fractions and could best be seen on rural or high altitude sites without pollution (figure 1).



**Figure 1:** NO<sub>2</sub> TotCol measured with Pandora 121 at Izaña on 13 Apr 2019 with r-codes nvs1 and nvs2.

NO<sub>2</sub> is a spatially heterogeneous trace gases and therefore the user should read section 6.1 before using nvs1 data.

r-code	nvs1
Code creator	Manuel Gebetsberger, 4 Nov 2019
DQ limits creator	Manuel Gebetsberger, 4 Nov 2019
Output product	NO2 TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.7 and higher
Product status	disused
Observation mode	Direct sun
Filters used	OPEN
Effective height	7.2 km
Number of fitting windows	1
Reference	Synthetic reference spectrum
Wavelength window	400.0 nm - 440.0 nm
Order of polynomials	SMO 4, OFFS 0, WLC 1, RSC -1
Fitted gases	O3-1 (225.0 K), NO2-1 (254.5 K)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	MEAS
AMF limits	7, 14
AtmVar limits	20, 40
Wavelength shift limits	0.05, 0.1
wrms limits	9.3e-4, 1.95e-3

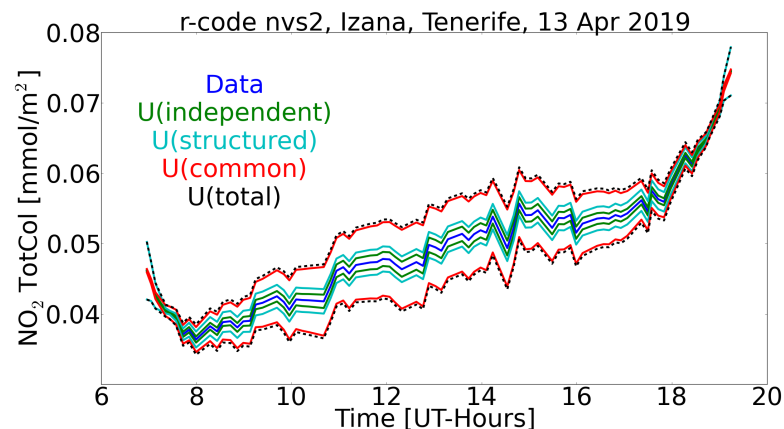


## 2.9 nvs2

nvs2 is the first code for NO<sub>2</sub> TotCols, which makes use of the new capabilities of the v1.8 processor. It has replaced nvs1 as official PGN total column NO<sub>2</sub> product code on 20 Nov 2020, and was itself replaced by nvs3 on 21 May 2021.

Here the main improvements of nvs2 over nvs1:

- Two spectral fittings are applied, which differ only by the EffTemp of NO<sub>2</sub>. It is taken from the boundary layer climatology in one case ("BL-clim") and from the stratospheric NO<sub>2</sub> climatology in the other case ("NO2s-clim"). Using the NO<sub>2</sub> stratospheric column climatology, a so-called tropospheric fraction is determined (see ATDB, section "Case 4").
- For the final NO<sub>2</sub> TotCol an effective height from the boundary layer climatology is applied to the tropospheric fraction and an effective height from the stratospheric NO<sub>2</sub> climatology is applied to the stratospheric fraction.



**Figure 2:** NO<sub>2</sub> TotCol measured with Pandora 121 at Izaña on 13 Apr 2019 with r-code nvs2 showing the different uncertainty outputs as developed for ProcVers 1.8.

With these changes we believe the major shortcomings of previous retrieval codes for NO<sub>2</sub> TotCols are significantly reduced. NO<sub>2</sub> is a spatially heterogeneous trace gases and therefore the user should read section 6.1 before using nvs2 data.

r-code	nvs2
Code creator	Alexander Cede & Martin Tiefengraber, 20 Nov 2020
DQ limits creator	Manuel Gebetsberger, 4 Nov 2019
Output product	NO2 TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	disused
Observation mode	Direct sun
Filters used	OPEN
Effective heights	BL+NO2s
Number of fitting windows	2
Reference	Synthetic reference spectrum
Wavelength window	400.0 nm - 470.0 nm
Order of polynomials	SMO 4, OFFS 0, WLC 0, RSC -1
Fitted gases	O3-2 (O3-clim), NO2-1 (BL-clim), O2O2-1 (O2O2-clim), H2O-1 (BL-clim), OIO-1 (BL-clim), I2-1 (BL-clim); O3-2 (O3-clim), NO2-1 (NO2s-clim), O2O2-1 (O2O2-clim), H2O-1 (BL-clim), OIO-1 (BL-clim), I2-1 (BL-clim)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	INSTR
AMF limits	7, 14
AtmVar limits	20, 40
Wavelength shift limits	0.05, 0.1
wrms limits	9.3e-4, 1.95e-3



## 2.10 nvs3

nvs3 is as nvs2, but uses updated quality flag thresholds based on *Gebetsberger et al.* [4]. It has replaced nvs2 as official PGN total column NO<sub>2</sub> product code on 21 May 2021.

r-code	nvs3
Code creator	Alexander Cede & Martin Tiefengraber, 20 Nov 2020
DQ limits creator	Manuel Gebetsberger, 21 May 2021
Output product	NO2 TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	official
Observation mode	Direct sun
Filters used	OPEN
Effective heights	BL+NO2s
Number of fitting windows	2
Reference	Synthetic reference spectrum
Wavelength window	400.0 nm - 470.0 nm
Order of polynomials	SMO 4, OFFS 0, WLC 0, RSC 0
Fitted gases	O3-2 (O3-clim), NO2-1 (BL-clim), O2O2-1 (O2O2-clim), H2O-1 (BL-clim), OIO-1 (BL-clim), I2-1 (BL-clim); O3-2 (O3-clim), NO2-1 (NO2s-clim), O2O2-1 (O2O2-clim), H2O-1 (BL-clim), OIO-1 (BL-clim), I2-1 (BL-clim)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	INSTR
AMF limits	7, 14
AtmVar limits	36, 42
Wavelength shift limits	0.05, 0.1
wrms limits	9.7e-4, 1.52e-3



## 2.11 ous1

ous1 is the first official PGN-code fitting the O<sub>3</sub> EffTemp, which is possible by using the synthetic reference spectrum instead of the theoretical one.

Although O<sub>3</sub> TotCols from ous1 are more accurate than from the simultaneously used code out2 due to the temperature fitting, out2 is also offered officially for the following two reasons:

- out2 is purely based on laboratory calibration, while ous1 needs field calibration. Hence ous1 cannot be used until such calibration is done. Furthermore, out2 is needed to produce a first guess O<sub>3</sub> TotCol for the field calibration of ous1.
- ous1 cannot be applied to all PGN instruments yet, since it requires absolute calibration and high-quality stray light calibration to be performed on a Pandora. For the instrument lacking such calibration, out2 can be used. Note that data based on the processor version 1.9 are spectral stray light corrected. For these datasets, ous1 will be included.

Please also note the general recommendations for data products retrieved in the UV in section 6.2.

r-code	ous1
Code creator	Martin Tiefengraber, 13 Aug 2020
DQ limits creator	Manuel Gebetsberger, 21 May 2021
Output product	O3 TotCol [mol/m <sup>2</sup> ], O3 EffTemp [K]
Processor requirement	1.9 and higher
Product status	official, official
Observation mode	Direct sun
Filters used	U340
Effective height	O3
Number of fitting windows	1
Reference	Synthetic reference spectrum
Wavelength window	305.0 nm - 333.0 nm
Order of polynomials	SMO 1, OFFS 0, WLC 0, RSC 0
Fitted gases	O3-2 (T-fit), NO2-1 (BL-clim), SO2-2 (BL-clim), HCHO-1 (BL-clim), HONO-1 (BL-clim), BrO-1 (O3-clim)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	INSTR
AMF limits	4, 7
AtmVar limits	38, 90
Wavelength shift limits	0.05, 0.1
wrms limits	8.7e-4, 1.29e-3



## 2.12 out0

out0 is the first code for O<sub>3</sub> TotCols provided within the PGN. However, due to its shortcomings described below, it was never officially used to perform satellite validation and is now disused and replaced by codes out2 and ous1.

All of the existing data with this code were processed with BlickP version 1.7, which reports the O<sub>3</sub> TotCol in DU, where 1 DU corresponds to 4.4615e-4 mol/m<sup>2</sup>. The total uncertainty for out0 data with DQF 0, which were processed with BlickP version 1.7, is typically estimated to 12 DU = 5.4e-3 mol/m<sup>2</sup>, based on [32].

The main shortcomings of out0 data and reasons for the limited accuracy, even if processed with v1.8, are listed below. They are discussed in more detail e.g. in Müller *et al.* [15] and Tiefengraber and Cede [26].

- The retrieval is purely based on laboratory calibration. Any non-smooth structure in the instrument's spectral sensitivity inside the fitting wavelength range (310 to 330 nm) can cause a systematic bias in the data. This is especially true for units, which have not undergone absolute calibration in the laboratory. Studies from year 2020 have shown, that this issue can be reduced with different settings for the spectral fitting, and retrieval code out1 has been developed as a consequence.
- The algorithm assumes a fixed effective O<sub>3</sub> temperature of 225 K, which commonly leads to underestimation (overestimation) of the total O<sub>3</sub> column in summer (winter). Therefore a seasonal difference between PGN data and other measurements, which take the EffTemp into account, is seen. This issue is reduced with code out1, which uses climatological data for the O<sub>3</sub> EffTemp and even more with code ous0, where the O<sub>3</sub> EffTemp is fitted.

out0 is retrieved in the short UV wavelength and therefore the user should read section 6.2 before using out0 data.

r-code	out0
Code creator	Alexander Cede, 20 Jan 2017
DQ limits creator	Alexander Cede, 20 Jan 2017
Data products	O3 TotCol [mol/m <sup>2</sup> ], SO2 TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.7 and higher
Product status	disused, unusable
Observation mode	Direct sun
Filters used	U340
Effective heights	20.4 km, 4.0 km
Number of fitting windows	1
Reference	Theoretical reference spectrum
Wavelength window	310.0 nm - 330.0 nm
Order of polynomials	SMO 4, OFFS 0, WLC 1, RSC -1
Fitted gases	O3-1 (225.0 K), NO2-1 (254.5 K), SO2-1 (259.2 K), HCHO-1 (256.9 K)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	MEAS
AMF limits	5, 7
AtmVar limits	20, 40
Wavelength shift limits	0.5, 1.0
wrms limits	0.01, 0.02



## 2.13 out1

out1 is the first code for O<sub>3</sub> TotCols, which makes use of the new capabilities of the v1.8 processor. As out0, out1 has also never been used as an official code and is now disused.

Here the main characteristics of out1 and improvements over out0:

- As for out0, the retrieval is purely based on laboratory calibration. However the sensitivity to non-smooth structures in the instrument's spectral sensitivity causing a systematic bias in the data has been greatly reduced by applying several changes to the fitting window: the wavelength range has been shifted down by 5 nm, only O<sub>3</sub> and SO<sub>2</sub> are fitted, and the orders of both the smoothing and offset polynomials have been reduced.
- Instead of using a fixed effective O<sub>3</sub> temperature, out1 uses the climatological value. This reduces the bias due to a mismatch of the true EffTemp and the one used in the retrieval significantly.

out1 is retrieved in the short UV wavelength and therefore the user should read section 6.2 before using out1 data.

r-code	out1
Code creator	Martin Tiefengraber, 3 Aug 2020
DQ limits creator	Alexander Cede, 20 Jan 2017
Output product	O3 TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	disused
Observation mode	Direct sun
Filters used	U340
Effective height	O3
Number of fitting windows	1
Reference	Theoretical reference spectrum
Wavelength window	305.0 nm - 325.0 nm
Order of polynomials	SMO 3, OFFS -1, WLC 0, RSC -1
Fitted gases	O3-2 (O3-clim), SO2-2 (BL-clim)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	INSTR
AMF limits	5, 7
AtmVar limits	20, 40
Wavelength shift limits	0.5, 1.0
wrms limits	0.01, 0.02

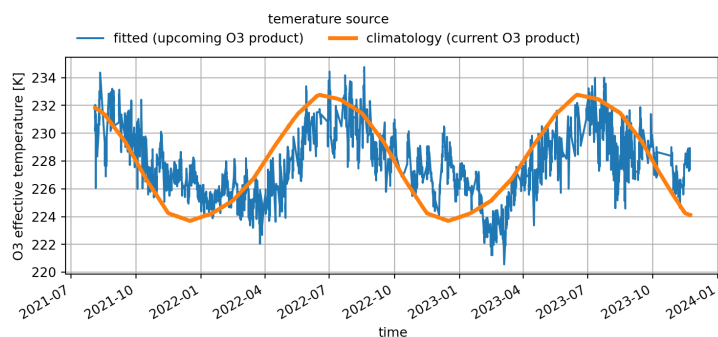




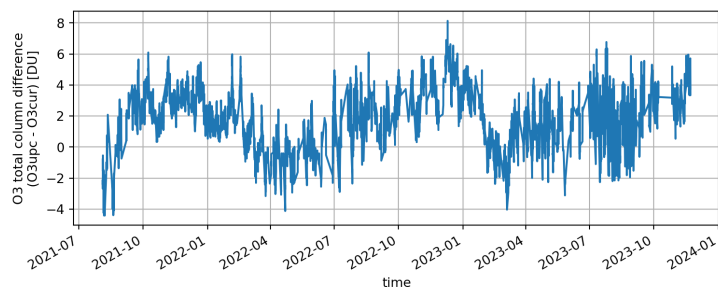
## 2.14 out2

out2 is as out1, but uses updated quality flag thresholds based on *Gebetsberger et al.* [4]. It is the first official PGN code for O<sub>3</sub> TotCols together with code ous1 (which will be distributed starting with the processor version 1.9).

Please be aware that offsets from the climatological effective O<sub>3</sub> temperature to the actual temperature will cause an apparent seasonal biases in O<sub>3</sub> TotCol. This is exemplarily shown in Figure 3 for the O<sub>3</sub> timeseries in Tsukuba, Japan (PI: Tamaki Fujinawa). The shifted climatological temperature (orange line in 3a) compared to the retrieved (from ous1, blue line in 3a) translates into an artificial seasonal bias in O<sub>3</sub> TotCol (3b, showing ous1 minus out2).



(a) O<sub>3</sub> effective temperature from the out2 (orange), where the O<sub>3</sub> temperature is taken from climatology and ous1 (blue), where the temperature is fitted.



(b) Total O<sub>3</sub> difference between out2 (=O<sub>3</sub>cur: effective temperature from climatology) ous1 (=O<sub>3</sub>upc: fitted effective temperature).

**Figure 3**

Please also note the general recommendations for data products retrieved in the UV in section 6.2.

r-code	out2
Code creator	Martin Tiefengraber, 3 Aug 2020
DQ limits creator	Manuel Gebetsberger, 21 May 2021
Output product	O <sub>3</sub> TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	official
Observation mode	Direct sun
Filters used	U340
Effective height	O <sub>3</sub>
Number of fitting windows	1
Reference	Theoretical reference spectrum
Wavelength window	305.0 nm - 325.0 nm
Order of polynomials	SMO 3, OFFS -1, WLC 0, RSC -1
Fitted gases	O <sub>3</sub> -2 (O <sub>3</sub> -clim), SO <sub>2</sub> -2 (BL-clim)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	INSTR
AMF limits	5, 7
AtmVar limits	64, 73
Wavelength shift limits	0.5, 1.0
wrms limits	6.40e-3, 6.76e-3



## 2.15 sus1

sus1 is the first official PGN code for SO<sub>2</sub> TotCols. It is similar to ous1, but with two major changes:

- The wavelength range is shifted down by 4 nm in order to grab an additional strong SO<sub>2</sub> absorption feature.
- The order of the smoothing polynomial is increased to 4 in order to reduce the influence of a faulty stray light correction on the results.

sus1 is retrieved in the short UV wavelength and with this processor version no spectral stray light correction is applied to the spectra. Furthermore, although a potential resolution change is taken into account for each measured spectrum in the spectral fitting (see table), this is not done for the (Fraunhofer) reference spectrum itself. As a consequence of both, an artificial seasonality can be observed (amplitude < 0.05 DU) in data sets where the reference spectrum had to be selected from days with relatively high solar zenith angles (e.g. winter in the northern hemisphere).

Please also note the general recommendations for data products retrieved in the UV in section 6.2.

r-code	sus2
Code creator	Martin Tiefengraber, 30 Oct 2020
DQ limits creator	Manuel Gebetsberger, 21 May 2021
Output product	SO2 TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	official
Observation mode	Direct sun
Filters used	U340
Effective height	BL
Number of fitting windows	1
Reference	Synthetic reference spectrum
Wavelength window	306.0 nm - 326.4 nm
Order of polynomials	SMO 4, OFFS 0, WLC 0, RSC 0
Fitted gases	O3-2 (T-fit), NO2-1 (BL-clim), SO2-2 (BL-clim), HCHO-1 (BL-clim), HONO-1 (BL-clim), BrO-1 (O3-clim)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	INSTR
AMF limits	3, 5
AtmVar limits	20, 40
Wavelength shift limits	0.05, 0.1
wrms limits	1e-3, 3e-3



## 2.16 wvt1

wvt1 is the first official direct sun H<sub>2</sub>O Totcol product of the PGN. While H<sub>2</sub>O offers a large number of potential fitting windows, wvt1 utilizes the strongest band still covered by both the UV and the VIS Pandora channels, and exhibits the least temperature dependence at the same time. This band is centered around 505 nm and features optical depths for standard abundances (1 cm) well above 1e-3. The determination of the quality flag thresholds was done according to *Gebetsberger et al.* [4].

<b>r-code</b>	<b>wvt1</b>
Code creator	Manuel Gebetsberger, 23 June 2022
DQ limits creator	Manuel Gebetsberger, 31 May 2023
Output product	H2O TotCol [mol/m <sup>2</sup> ]
Processor requirement	1.8 and higher
Product status	official
Observation mode	Direct sun
Filters used	OPEN
Effective height	BL
Number of fitting windows	1
Reference	Theoretical reference spectrum
Wavelength window	492.0 nm - 510.0 nm
Order of polynomials	SMO 1, OFFS -1, WLC 0, RSC -1
Fitted gases	H2O-1 (BL-clim), O3-2 (O3-clim), NO2-1 (BL-clim)
Ring	Not fitted
Molecular scattering	Subtracted
Uncertainty	INSTR
AMF limits	5, 7
AtmVar limits	40, 70
Wavelength shift limits	0.5, 1.0
wrms limits	3.78e-3, 4.59e-3



### 3 Cross section sources

Table 2 lists the sources for the trace gas cross sections used in the spectral fitting. Column "Name" combines the acronym of the trace gas (e.g. "BrO") with an index to distinguish different sources used.

**Table 2:** Cross section sources

Name	Source
BrO-1	<i>Wilmouth et al.</i> [31]
HCHO-1	<i>Meller and Moortgat</i> [14]
HONO-1	<i>Stutz et al.</i> [23]
H2O-1	HITRAN 2012 <i>Rothman et al.</i> [18]
I2-1	<i>Saiz-Lopez et al.</i> [19]
NO2-1	<i>Vandaele et al.</i> [28]
OIO-1	<i>Spietz</i> [21]
O2O2-1	<i>Thalman and Volkamer</i> [24] for $\lambda \leq 596$ nm combined with <i>Newnham and Ballard</i> [16] for $\lambda > 596$ nm
O3-1	<i>Serdyuchenko et al.</i> [20] for $\lambda \leq 1100$ nm combined with <i>Rothman et al.</i> [18] for $\lambda > 1100$ nm
O3-2	Like O3-1, but the wavelength registration of the <i>Serdyuchenko et al.</i> [20] data is shifted by +3 pm
SO2-1	<i>Vandaele et al.</i> [29]
SO2-2	<i>Vandaele et al.</i> [30]



## 4 Viewing geometries

This section gives a short explanation of the air mass sampled in the different observation modes. It uses the "effective height" ( $h_{EFF}$ ) of a trace gas, which is given by:

$$h_{EFFj} = \frac{\int_{SURF}^{ToA} n_j(h) \cdot h \cdot dh}{\int_{SURF}^{ToA} n_j(h) \cdot dh} \quad (1)$$

The integral runs from the surface SURF to the top of the atmosphere ToA along the vertical path  $h$ .  $n_j(h)$  is the particle density of the trace gas at height  $h$ .

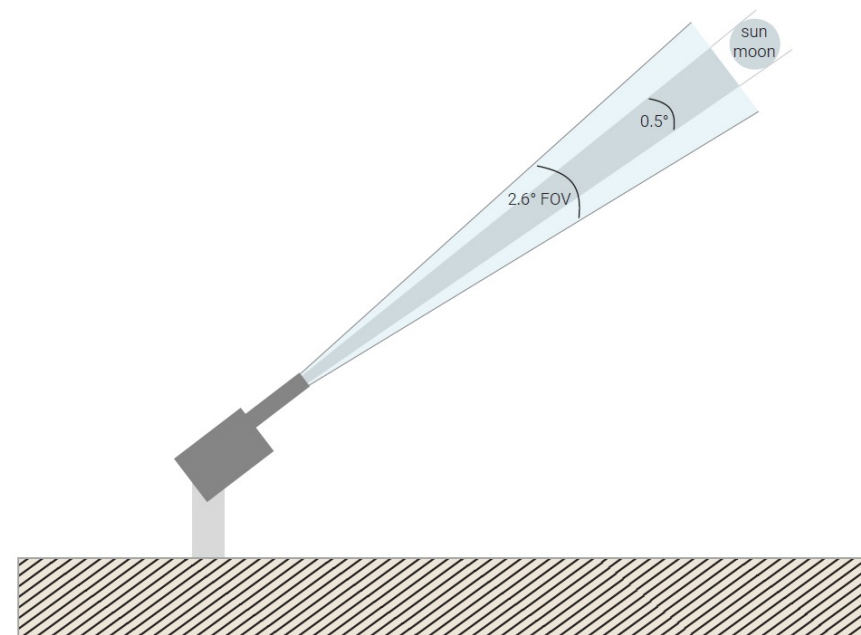
Another expression introduced is the "Effective Ground Location" (EGL). The EGL is defined as that location on the ground, for which the measurements are "most representative", which is in general NOT the location of the ground-based instrument or the center of the satellite's footprint projected to the ground.

### 4.1 Direct sun or moon

The viewing geometry of the direct sun or moon observation mode is outlined in figure 4. The sampled air mass is a circular cone with its apex at the entrance of the instrument and extending into the direction of the sun or moon. This means that for direct sun observations, the measurements sample air towards East in the morning, South around noon, and West in the afternoon (for Northern Hemispheric locations).

For PGN direct sun observations, the opening angle of the cone is about  $\alpha = 2.6^\circ = 0.045$  rad, which is the FWHM angle of the Pandora FOV in direct sun mode. For direct moon observations, the FWHM of the FOV is  $\alpha = 1.5^\circ = 0.026$  rad. However the vast majority of the light sampled in direct observation mode comes from an angle of  $0.5^\circ = 0.009$  rad, since this is the angular size of both the sun and the moon. The direct light comes from inside these  $0.5^\circ$ , the forward scattered diffuse light is distributed over the entire FOV of the instrument. Only at large ZA and high aerosol content is the diffuse fraction significant (a few percent of the total signal) and can possibly introduce systematic errors in the retrieval. Each trace gas

molecule inside the cone is equally "counted", i.e. there is no dependence of the vertical profile in the data.



**Figure 4:** Direct sun observations

Figure 5 shows a simplification for the situation for direct sun observations.  $R$  is the distance from the center of the Earth to the measurement location (about 6370 km, refined based on the location's latitude),  $ZA^*$  is the apparent solar (or lunar) zenith angle (i.e. the geometrical ZA corrected for refraction) and  $h_{EFF}$  is the effective height for the trace gas (from equation 1). Note that if  $h_{EFFj}$  is smaller than 10 km ("tropospheric effective height"), the effective height is added to the station height. Otherwise ("stratospheric effective height"), the station height is ignored and the effective height is added to the Earth's surface.  $s$  is the slant distance between the instrument and the point where the direct beam is at height  $h_{EFF}$  in the atmosphere.  $d$  is the ground distance between the location of the instrument and location underneath the point where the direct beam is at height  $h_{EFF}$  in the atmosphere.  $ZA'$  is the "reduced" zenith angle, which can be calculated by equation 2.



$$ZA' = \arcsin \left[ \left( \frac{R}{R + h_{\text{EFF}}} \right) \cdot \sin(ZA^*) \right] \quad (2)$$

s and d can be calculated with equations 3 and 4 respectively.

$$s = R \cdot \frac{\sin(ZA^* - ZA')}{\sin(ZA')} \quad (3)$$

$$d = R \cdot (ZA^* - ZA') \quad (4)$$

In combination with the solar or lunar azimuth, d can be used to estimate the EGL for direct observations. The estimation of the direct AMF in the PGN retrievals, used to convert the measured slant columns into vertical columns, is based on some assumptions, e.g. that the vertical distribution of the trace gas is a delta function at  $h_{\text{EFF}}$ . It is given by equation 5. For details on this equation see e.g. *Bernhard et al.* [8].

$$\text{AMF}_{\text{DIR}}(ZA^*) = \sec(ZA') \quad (5)$$

The accuracy of  $\text{AMF}_{\text{DIR}}$  depends mostly on how well  $h_{\text{EFF}}$  represents the "truth", but is in general better than 2% for  $ZA < 80^\circ$ .

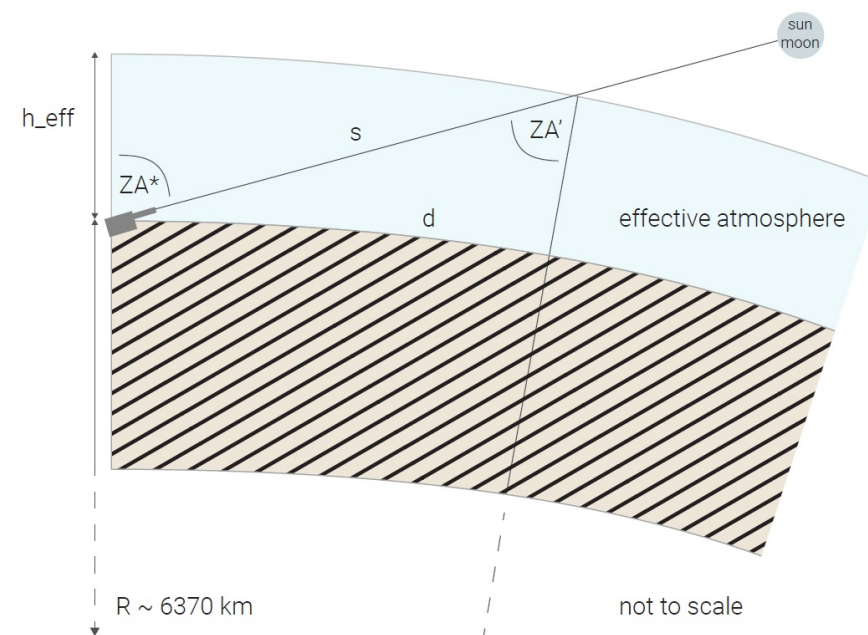


Figure 5: Direct sun geometry

## 4.2 Profile

The viewing geometry of profile measurements ("MAXDOAS") is outlined in figure 6. The Pandora FOV for sky observations is circular with FWHM of  $\alpha = 1.5^\circ = 0.026$  rad. The photons captured by the instrument originate from the sun, enter the atmosphere in the direction of the solar  $ZA$ , and then make one or more interactions to be directed into the instrument. Some of them are reflected by the surface, others are scattered in the atmosphere to end up in the instrument's entrance optics. A smaller fraction does even multiple interactions, e.g. reflection plus scattering, or two scatter processes, etc. Therefore the sampled air mass is a rather large "region" above the instrument. The EGL depends on where the instrument is exactly pointing and therefore changes for each of the elevation angles sampled by the instrument. Hence for profile observations, the EGL of the measurement is usually shifted relative to the instrument's location, namely into the direction of the pointing azimuth.

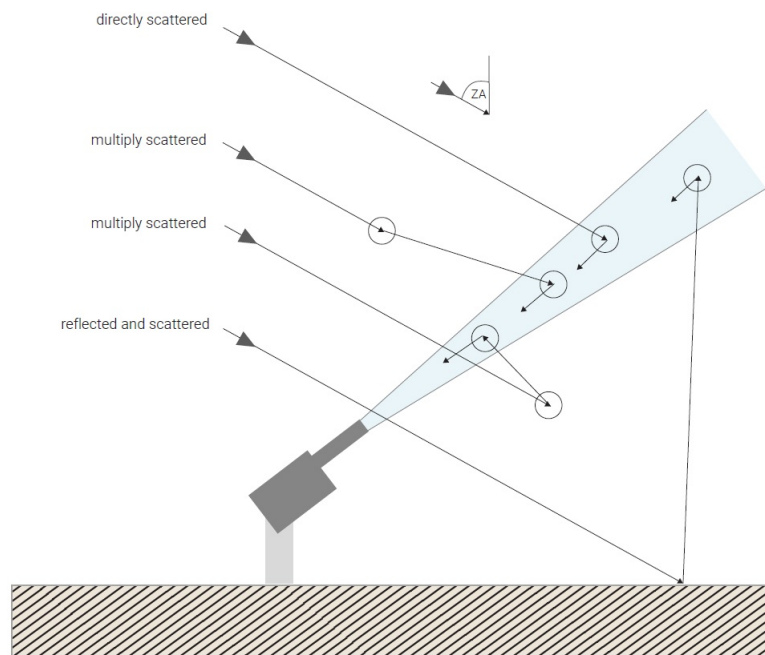


Figure 6: Profile observations

### 4.3 Satellite nadir view

The viewing geometry of near nadir observations from satellite (or aircraft) is outlined in figure 7. The FOV of the instrument projected to the ground gives the so-called footprint of the satellite instrument. In current satellites, the diameters of the footprint vary from about 5 km to more than 100 km. As for profile measurements, the photons captured by the satellite instrument originate from the sun, enter the atmosphere in the direction of the solar ZA, and then make one or more interactions to be directed into the satellite instrument. The sampled air mass is a rather large "region" underneath the satellite and in the direction of the Sun. Hence for nadir observations, the EGL is also shifted relative to the center of the footprint, namely towards East in the morning, South around noon, and West in the afternoon (for Northern Hemispheric locations). However the "displacement" into the direction of the sun is not as large as for direct observations.

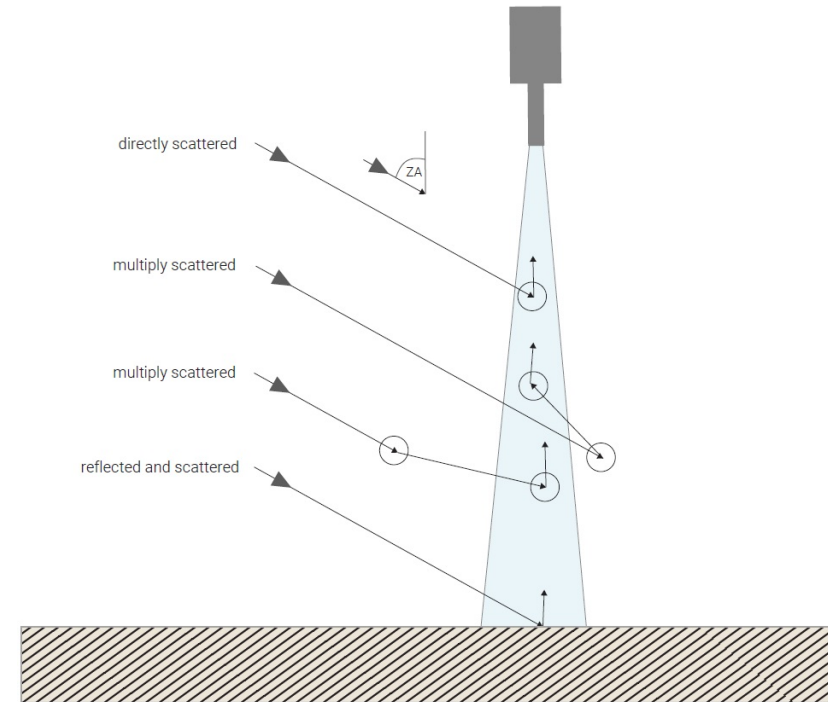


Figure 7: Nadir observations



## 5 Uncertainty and data quality information

This section is a summary of section "Data uncertainty" in the ATBD. The basic principle of the uncertainty and data quality information in the PGN data products is this:

- Those error sources, which we can quantitatively propagate through the retrieval algorithm, are included in the output uncertainty of the L2 data (section 5.1).
- Those error sources, which we know exist, but we cannot quantify, are either included in the DQF of the L2 data (section 5.3), or not included at all (section 5.4).

Hence to distinguish between "good" and "bad" data, both the uncertainty output and the DQF should be used. As already mentioned in section 2, only the L2 data in observation mode have a complete set of uncertainties as described in section 5.1. Those measured in observation mode Profile have a more limited uncertainty information.

### 5.1 Uncertainty types

Three types of uncertainties are distinguished, which differ from each other by the correlation length in time:

- Independent uncertainty  $U_I$ : the correlation length in time is zero. An example is the photon noise propagated into the TotCol measured at a certain moment, which is totally uncorrelated to the propagated photon noise for measurements taken at any other time.
- Common uncertainty  $U_C$ : the correlation length in time is infinite. An example is an error in the assumed slant column in the reference spectrum. This error affects all retrieved 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.

- Structured uncertainty  $U_S$ : the correlation length in time is larger than zero, but not infinite. An example 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 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 nearly 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.

For the total uncertainty  $U$  of a single data point, we simply combine  $U_I$ ,  $U_C$  and  $U_S$  as shown in equation 6:

$$U = \sqrt{U_I^2 + U_C^2 + U_S^2} \quad (6)$$

When the data are averaged, e.g. by building the mean TotCol 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  $U(n,short)$ , is given by:

$$U(n,short) = \frac{1}{n} \cdot \sqrt{\sum_{i=1}^n [U_I(i)^2] + \left[ \sum_{i=1}^n U_C(i) \right]^2 + \left[ \sum_{i=1}^n U_S(i) \right]^2} \quad (7)$$





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 TotCol 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.

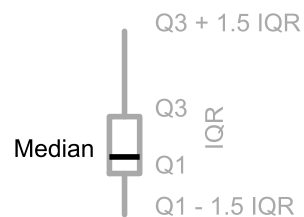
The other extreme case assumes the structured uncertainties to be uncorrelated along the dimension. In this "long" case the total uncertainty  $U(n, \text{long})$ , is given by:

$$U(n, \text{long}) = \frac{1}{n} \cdot \sqrt{\sum_{i=1}^n [U_I(i)^2] + \left[ \sum_{i=1}^n U_C(i) \right]^2 + \sum_{i=1}^n [U_S(i)^2]} \quad (8)$$

Here the structured uncertainty "behaves" like the independent uncertainty. An example for this would be the mean TotCol 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.

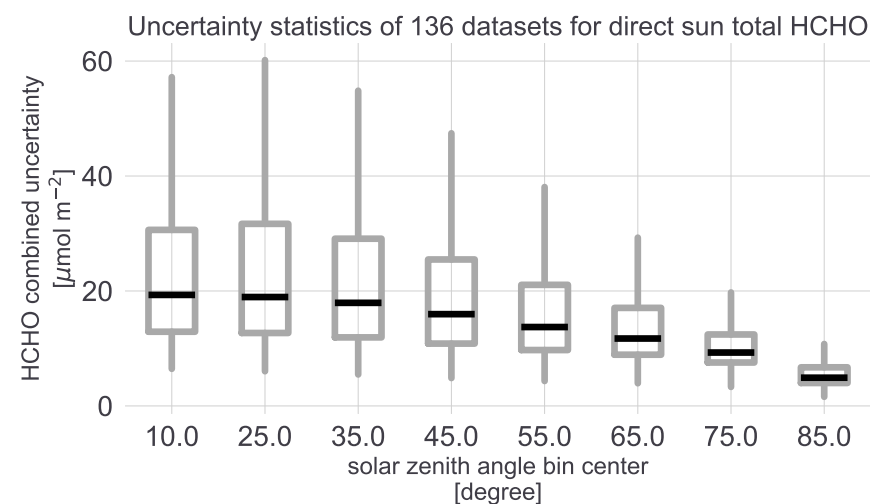
## 5.2 Total uncertainty for operational direct sun products

Based on the available PGN data from January 16<sup>th</sup> 2025, we have calculated the total uncertainty (equation 6) for all operational products at that time. As the SZA has a strong influence on the total uncertainty, we have binned the data accordingly, considering only high quality data. Each SZA bin is summarized as a box-plot with the box and whiskers as defined in the legend on the right (Q1=1<sup>st</sup> quartile (25%), IQR=inter-quantile-range). Table 3 provides a guide to the individual plots.

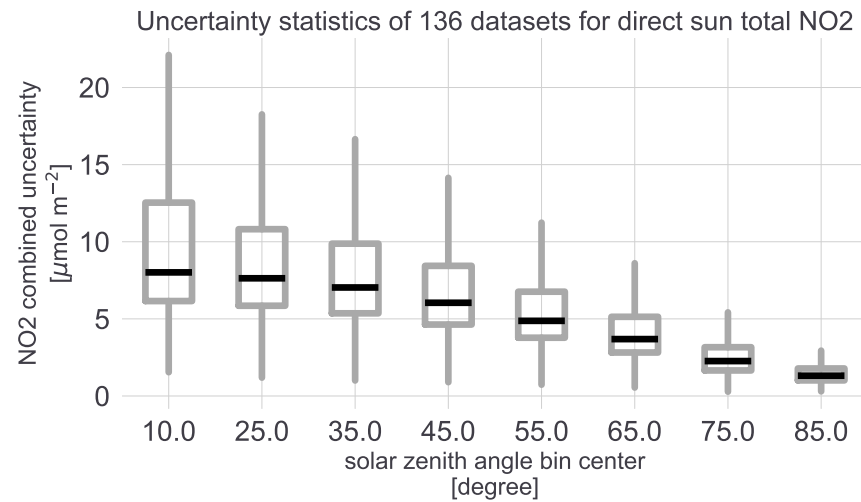


**Table 3**

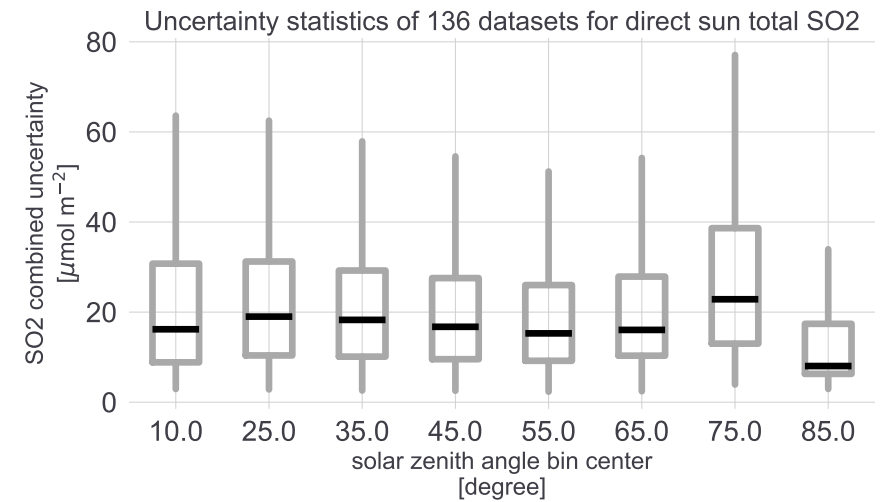
Product	Version	Figure	Data unit	Datasets
HCHO TotCol	rfus5p1-8	8	$\mu\text{mol m}^{-2}$	136
NO2 TotCol	rnvs4p1-8	9	$\mu\text{mol m}^{-2}$	136
O3 TotCol	rout2p1-8	10	$\text{mmol m}^{-2}$	172
SO2 TotCol	rsus1p1-8	11	$\mu\text{mol m}^{-2}$	136
H2O TotCol	rwvt1p1-8	12	$\text{mol m}^{-2}$	172



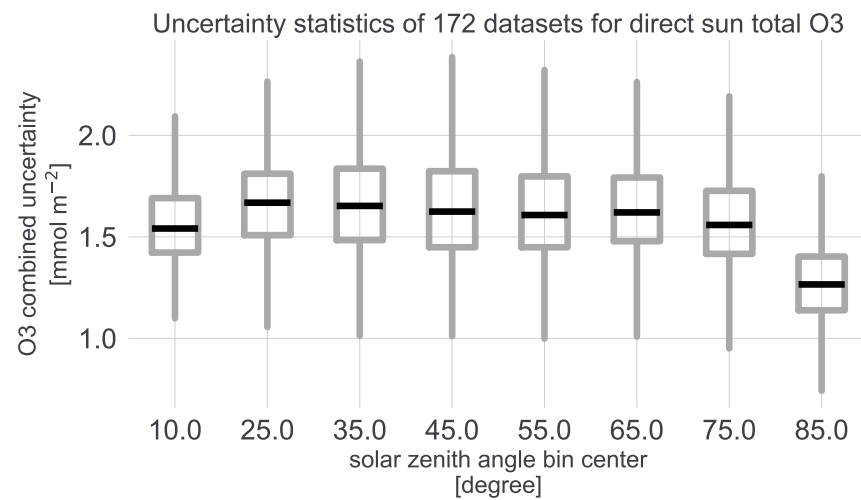
**Figure 8:** Distribution of the total uncertainty of HCHO TotCol as a function of solar zenith angle. The center line marks the median value, the box spans the inter-quantile-range (IQR) and the whiskers extend up to +/- 1.5 of the IQR. Only high quality data are included.



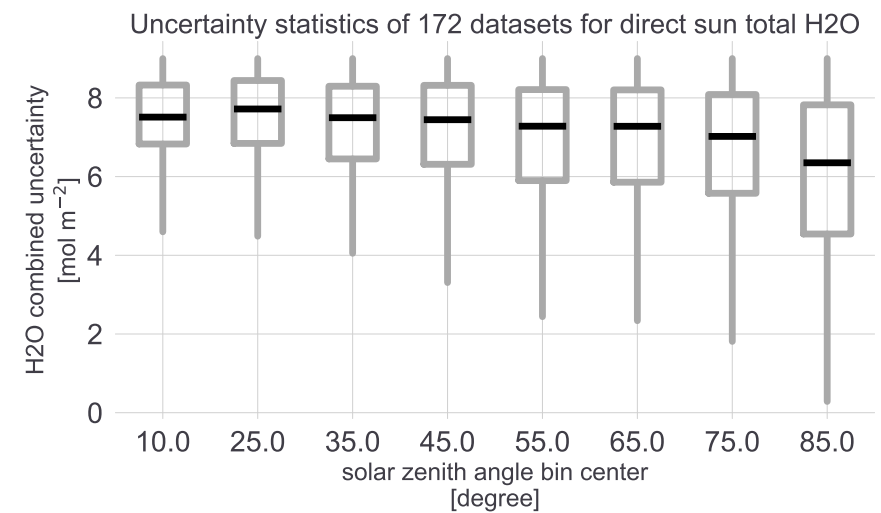
**Figure 9:** Explanation as for Figure 8, but for NO<sub>2</sub> TotCol.



**Figure 11:** Explanation as for Figure 8, but for SO<sub>2</sub> TotCol.



**Figure 10:** Explanation as for Figure 8, but for O<sub>3</sub> TotCol.



**Figure 12:** Explanation as for Figure 8, but for H<sub>2</sub>O TotCol.



## 5.3 Data Quality Flags

Currently there are 9 possible DQFs: DQ0, DQ1, DQ2, DQ10, DQ11, DQ12, DQ20, DQ21 and DQ22. The unit position 0, 1 or 2 has the following meaning:

- Unit position 0, "High quality": no data quality indicator exceeds the data quality 1 (DQ1) limit and therefore there are no indications that the data might not be of the highest possible quality. Once quality assured, those data can be used with high confidence and the uncertainty associated with these additional effects is negligible, i.e. the provided uncertainty can be assumed to represent the true uncertainty.
- Unit position 1, "Medium quality": at least one data quality indicator exceeds the DQ1 limit and therefore the quality of the data might be reduced. Depending on the application, the user should decide whether to use these data. Note that the reduced quality can originate from instrumental sources (e.g. too large wavelength shift) or atmospheric sources (e.g. clouds in direct sun measurements). This also means that the reported data underestimates the true uncertainty by a small amount.
- Unit position 2, "Low quality": at least one data quality indicator exceeds the data quality 2 (DQ2) limit and therefore the quality of the data is strongly reduced. For most purposes, the user should not use these data. As for unit position 1, the low quality can originate from instrumental or atmospheric sources. In this case it is nearly certain that reported data uncertainty underestimates the true uncertainty.

The decade of the DQF can be 0, 1 or 2 and has the following meaning:

- Decade 0, "Quality assured": this means that quality control (QC) has already been applied to the data. QC consists of a semi-automatic inspection of the obtained L2 data to determine, whether the instrument was monitoring correctly over a certain measurement period. With this procedure one can possibly detect factors influencing the data quality, which are not reflected in the uncertainty and not captured by the DQF. An example would be a dirty entrance window. This reduces the overall throughput of the system and is interpreted by the algorithm as an increased aerosol load in the atmosphere.

Since this effect causes usually a non-physical dependence of the AOD on the solar ZA, it can often be detected in the QA process. After the QC has been performed, the unit position will not change anymore.

- Decade 1, "Not quality assured": the quality flag in the unit position is based on an initial estimate and formal QC has not been carried out. It is possible that the unit position changes after that happens.
- Decade 2, "Unusable data": this means the data are not usable, which can have the following reasons:

The QC has revealed that there were data quality issues during that time.

r-code entry "Product status" says "unusable" for the respective output gas. This is typically the case for a trace gas that is not the "primary" gas to be fitted, for which the fitting setup, especially the wavelength range, is not optimized. These data will not undergo QC.

- The synthetic reference spectrum is used, but no calibration was applied to it. These data will not undergo QC.

It is also important to note, that the unit position of the DQF can only stay the same or increase through the different processing levels, but never go down.

## 5.4 Not covered by uncertainty or DQF

The following list describes effects, which are not included in the uncertainty output or the DQF, but we know that they can contribute to the data uncertainty. It is planned that these effects are included in the future. Obviously it is also possible that other effects, which are unknown to us and therefore not listed here, may contribute to the uncertainty.

- Cross sections: Neither the choice of the cross sections nor their uncertainty is currently taken into account.
- Algorithm: The spectral fitting algorithm 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 parameters. These effects are not considered in the current software version.



- Spectral fitting quality: Although the spectral fitting residual RMS is used as a threshold in the DQF, its impact on the uncertainty is not considered.
- Theoretical reference spectrum: Any systematic issues arising from using this reference not measured by the instrument itself are neglected.



## 6 Special notes

This section lists notes, which apply the specific groups of data products. The notes are referenced in the readme pages.

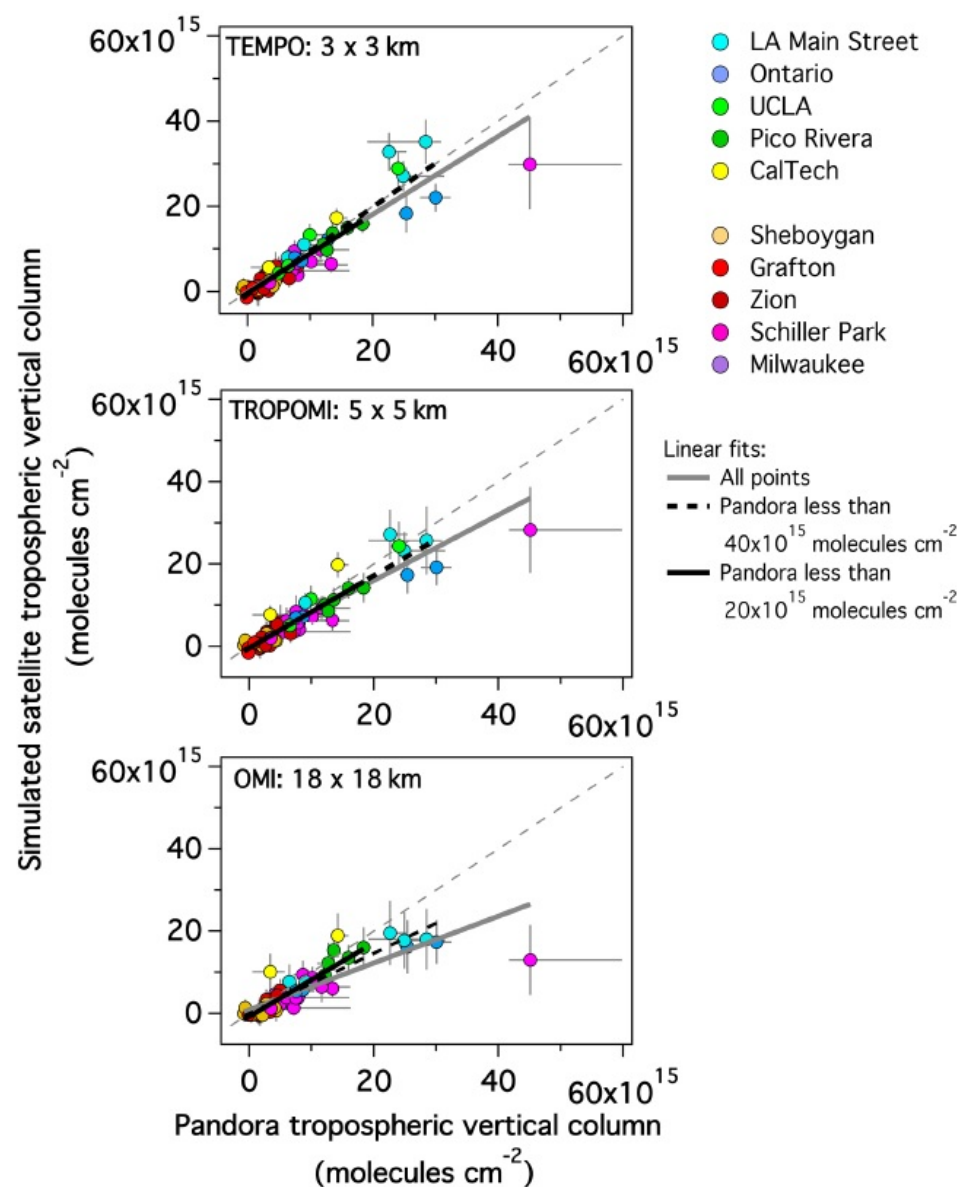
### 6.1 Note on spatially heterogeneous trace gases

Short-lived reactive trace gases with point sources on the surface are usually very heterogeneous both temporally and spatially. An example for this is  $\text{NO}_2$ , where the surface concentration and even column amounts can vary over a distance of meters and a time of seconds only.

For these species parameters such as the sampled air mass and the EGL of the measurements need to be considered when the data are compared to other measurements, e.g. surface in-situ data or satellite data (see also section 4). E.g. at solar ZA =  $60^\circ$  and an (true) effective height of the  $\text{NO}_2$  of 4.2 km, the effective location is shifted for a distance of 7.3 km in the direction of the sun relative to the location of the instrument (equation 4).

It turns out that for such gases the different viewing geometry between the PGN observations and satellite data is most often the main driver for possible differences between these products. The magnitude of this difference typically grows with the size of the satellite footprint and the pollution level at the location. An excellent description of this can be found in *Judd et al.* [13], from which figure 13 is taken.

Past studies have shown, that the satellite data being a spatial average over the footprint compare better to a temporal average from the ground measurements than to the instantaneous observation exactly at the satellite's overpass time (see e.g. *Cede et al.* [10]). Therefore it may be useful to build such a temporal average using the formula given in equation 7 when doing satellite to ground comparisons.



**Figure 13:** Figure 9 from *Judd et al.* [13]. Scatter plot of simulated satellite TropCols for different areal resolutions versus Pandora TropCols. The gray dashed line indicates the 1:1 line.



## 6.2 Note on retrievals in the short UV range

Pandora data retrieved in the short UV range (below 350 nm) are affected by spectral and spatial stray light. The aspects described below should be known when using PGN data products from this wavelength range.

- Spectral stray light: since the intensity of the solar light increases significantly from the UV into the visible spectral range, spectral stray light is affecting significantly the measurements in the short UV range. A dedicated correction is applied on the PGN spectra only for processor versions > 1.8. Hence a residual effect is likely, if no or just an insufficient stray light calibration is performed on an instrument. This causes e.g. total ozone columns to be biased low starting at solar ZA=70° deg and sometimes even at smaller values.
- Spatial stray light: the fraction of diffuse light entering the instrument in addition to the direct beam for direct sun or moon observations increases with the (solar or lunar) zenith angle and the amount of aerosols in the atmosphere. Since the direct algorithm is based on the assumption that only direct light is measured, a systematic error in the data is introduced, which grows with this "diffuse fraction". This causes e.g. total ozone columns to be biased low. However BlickP does currently not take this effect into account at all.

To "capture" this stray light problem, the AMF limits in such "short-UV" retrievals are set to lower values compared to retrieval at from higher wavelength regions.

## 6.3 Note on weak absorbers

We consider a trace gas a weak absorber, if the largest differential vertical optical depth is about 1e-3 or even below for typical column amounts. Such trace gases are more sensitive to measurement uncertainty than the more absorbing ones. This is described here, separately for direct sun observations and sky observations.

### 6.3.1 Weak absorbers in direct observation

Direct (sun or moon) observations have one major weakness with respect to spectral fitting compared to sky observations. The direct beam is quasi-parallel with a very

small divergence of 0.5°, while the sky radiance is rather homogeneous over the FOV of the system. Therefore the direct observation is extremely sensitive to the exact pointing of the instrument. When two direct measurements are made with a pointing difference relative to the source (e.g. the sun's direction) even below 0.1°, the spectra typically show a difference, which can significantly affect the column retrievals of weak absorbers. Here we call this spectral difference an "unwanted spectral signal" (USS). In order to minimize this effect, the Pandora spectrometer system uses a wedged entrance window and a diffuser in the filterwheel, when measuring in direct sun mode (see e.g. *Tiefengraber and Cede* [25]). Nevertheless the USS still exists, more for some instrument than for others, and we are still investigating options to further reduce it.

At this moment, BlickP cannot identify data affected by this USS, hence this does not show up in the uncertainties, not does it trigger some DQ thresholds. Therefore the user should be aware, that it is possible, that the retrieved columns amounts for weak absorbers have temporary systematic issues.

### 6.3.2 Weak absorbers in sky observation

While sky observations usually do not suffer from an USS as described in the previous section, they typically suffer from a low SNR. The amount of sky light reaching the instrument is simply not large enough to obtain a high SNR with a system like Pandora that uses an uncooled detector and small optics. Therefore the user should be aware that Profile retrievals for weak absorbers may be driven by the independent uncertainty (noise). In order to reduce the noise one can average the data as described in section 5.1.