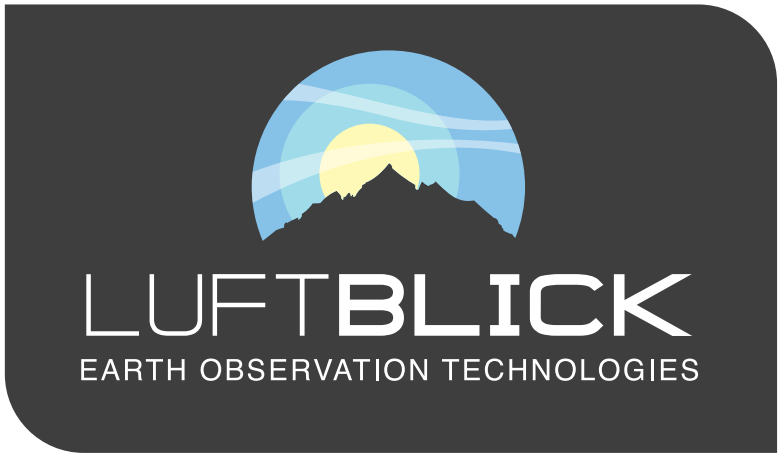


LuftBlick Report 2019005

Fiducial Reference Measurements for Air Quality

New Algorithm & Product Development Plan

Version 8.0, 31st Dec 2022

	Name	Company
prepared by	Martin Tiefengraber	LuftBlick
	Manuel Gebetsberger	LuftBlick
	Alexander Cede	LuftBlick
	Moritz Müller	LuftBlick
	Ragi Rajagopalan	LuftBlick

Contents

Document Change Record

Issue	Date	Section	Observations
1.0	30th Jun 2019	All	First version
2.0	14th Jan 2019	All, but in particular 3.5, 3.6, 3.7 and A	Major extensions in the direct sun total O_3 and SO_2 sections. Direct sun total HCHO section was updated too. Section calibration improvements was restructured and extended. Two appendix sections were added.
2.1	15th Jan 2020	All	Minor corrections and rewording.
3.0	28th Jun 2020	3.5, 3.7	Major extensions regarding direct sun total O_3, effective O_3 temperature. Direct sun total HCHO updates.
4.0	10th Jan 2021	All, but in particular 3.5.5, 3.6.1, 3.7.2, 4 and D have been included.	Section about data product validation was added. Additions to SO_2 and HCHO fitting setup. New appendix section about statistical intercomparison framework.
4.1	13th Jan 2021	4.3.1	Added O_3 temperature climatology to Figure 73
5.0	30th Jun 2021	3.2.2, 3.3.1, 3.10, 3.14.2, 3.18, 7	Next to minor additions to the stray light section and the ATBD, new data products are discussed: total column O_2O_2 retrievals and spectrometer 2 products for total column NO_2 (direct sun), NO_2 surface concentration and tropospheric column from sky observations. Finally, a discussion on the first stage improvements for spectral AOD was added.
5.1	2nd Jul 2021	Several	Minor additions for clarification.
6.0	31st Dec 2021	3.4, 3.8, 3.12, 3.13, 3.16, 3.18.2, 4.2, 4.7	Minor adjustments in the hole document, new data products are discussed: lunar total column NO_2, total column HONO, total column H_2O+validation, total column O_2. The second stage improvements for spectral AOD are discussed as well.
6.1	17th Jan 2022	3.12	Restructured description of fitting setup D for clarification.
7.0	28th Jun 2022	3.9, 3.12, 3.17 4.7	Direct sun total CHOCHO, H_2O, sky data profiles, H_2O validation
8.0	31st Dec 2022	3.8, 3.9, 5, 6,3.11,3.12	Minor additions to HONO and CHOCHO sections, H_2O from VIS spectrometer update, new section for direct sun total NO_3, new section on current product development limitations and final product release table have been added.

Acronyms and Abbreviations

$\lambda 1$	Fitting window starting wavelength
1σLEV	1σ uncertainty level
2σLEV	2σ uncertainty level
2σUNC	2σ-Uncertainty
NO_2	Nitrogen dioxide
NO_3	Nitrate radical
SO_2	Sulfur dioxide
AMF	Air mass factor
AOD	Aerosol Optical Depth
ATBD	Algorithm Theoretical Basis Document
AXC-C	Auto Cross Calibration for Column
AXC-T	Auto Cross Calibration for Temperature
BIRA	The Royal Belgian Institute for Space Aeronomy
BPol	Background polynomial order
BSS	Blick Software Suite
CHOCHO	Glyoxal
CrossComp	Comparison of direct sun and sky data retrievals of a Pandora.
DM	Direct moon
DS	Direct sun
DU	Dobson Units
E-MLE	Extended Minimum Langley Extrapolation
ESA	European Space Agency
FOV	Field of View
FRM4AQ	Fiducial Reference Measurements for Air Quality
FWHM	Full Width at Half maximum
HONO	Nitrous Acid
InterComp	Direct sun intercomparison of co-located Pandoras
L1	Level 1
LitRef	O_3 columns using a literature reference and O_3temp climatology
LSF	Line spread function
MAX	Multi axis
MLE	Minimum Langley Extrapolation
ND	Neutral density
O3temp	Effective ozone temperature
OD	Optical Depth
Pandora	Pandora spectrometer system
pm	Pico meter
POp	Pandonia Operations
PSS	Pan Software Suite
RMS	Root mean square of the spectral fitting residuals
Rome-SAP	University of Rome La Sapienza
S1	Spectrometer 1
S2	Spectrometer 2
SC	Slant column
SIMPLE	SIMulate Pandora data LEvels
SNR	Signal-to-noise-ratio
SpatSL	Spatial Stray Light
SpecSL	Spectral Stray Light
SyntRef	O_3 columns sing a synthetic reference and O_3temp fitted
SZA	Solar zenith angle
TN	Technical Note
USS	Unwanted spectral signal
UV	Ultraviolet

1 Introduction

This report is the combination of deliverables 4 (D4) and 11 (D11) of the ESA project “Fiducial Reference Measurements for Air Quality” (FRM4AQ) [1, 2]. In several instances this report refers to the parallelly running project Pandonia Operations (POp) [3, 4], since the work in both projects is complementary.

D11 is the Pandonia New Algorithm and Product Development Plan. Section 2 focuses on the evolution of the processing software, section 3 describes the new algorithms and products as planned in this project and section 6 gives an estimated schedule for implementation.

D4 is Algorithm Theoretical Basis Document (ATBD) of new/updated algorithms and technical notes (TN) and publications. It is described in section 7. Note that the ATBD itself is given in *Cede* [6]. The section just links to it.

In the appendix section three further topics are discussed. Section A talks about

implications for a O3 products using a literature reference in the retrieval, section B presents uncertainty estimations for a spectral AOD product and finally section D describes the statistical framework used for dataset comparison with validation activities.

Please note that in terms of units used in this report, you will find also Dobson Units (DU) beside standard SI units. This is due to the fact that in the previous data processor version(s), retrievals of column amounts of atmospheric constituents have been reported in DU.

1.1 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] Pandonia Operations [Proposal], LuftBlick Proposal 201804OPE, Issue 1, 2018.

[4] Pandonia Operations [Contract and Statement of Work], ESA Contract No. 4000124223/18/I-SBo, 2018.

[20] A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2018005: Final Pandonia report, 2018.

[6] A. Cede. *Manual for Blick Software Suite 1.8, Version 1-8-3*, 2021. URL `https://www.pandonia-global-network.org/wp-content/uploads/2021/05/BlickSoftwareSuite_Manual_v1-8-3.pdf`.

[21] A. Cede and M. Tiefengraber. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2014007: Pandora-2S Maintenance - Quality Assurance Plan, 2014.

[8] A. Cede, M. Tiefengraber, M. Gebetsberger, and E. Spinei Lind. Pandonia Global Network Data Products README Document, 2020.

[32] A. Kreuter. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2018003: Recommendations on operational AOD measurements for Pandonia, 2018.

[37] M. Müller, M. Tiefengraber, and A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2016011: Validation reports, 2016.

[47] 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.

[49] M. Tiefengraber and A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2017002: Network intercomparison campaign report, 2017.

1.2 Reference Documents

[13] Cospar international reference atmosphere: 1986 (0 km to 120 km). URL `http://ccmc.gsfc.nasa.gov/modelweb/atmos/cospar1.html`.

[14] R Atkinson, DL Baulch, RA Cox, JN Crowley, RF Hampson, RG Hynes, ME Jenkin, MJ Rossi, J Troe, and IUPAC Subcommittee. Evaluated kinetic and photochemical data for atmospheric chemistry: Volume ii–gas phase reactions of organic species. *Atmospheric chemistry and physics*, 6(11):3625–4055, 2006.

[15] S. Beirle, J. Lampel, C. Lerot, H. Sihler, and T. Wagner. Parameterizing the instrumental spectral response function and its changes by a super-gaussian and its derivatives. *Atmospheric Measurement Techniques*, 10(2): 581–598, 2017. doi: 10.5194/amt-10-581-2017. URL `https://amt.copernicus.org/articles/10/581/2017/`.

[16] P. K. Bhartia and C. Wellemeyer. *TOMS-V8 total O3 algorithm ATBD*, 2002.

[17] Samuel M Brohede, Craig S Haley, Chris A Mclinden, Christopher E Sioris, Donal P Murtagh, Svetlana V Petelina, Edward J Llewellyn, Ariane Bazureau, Florence Goutail, Cora E Randall, et al. Validation of odin/osiris stratospheric no2 profiles. *Journal of Geophysical Research: Atmospheres*, 112(D7), 2007.

[18] Claudia Rivera Cárdenas, Cesar Guarín, Wolfgang Stremme, Martina M. Friedrich, Alejandro Bezanilla, Diana Rivera Ramos, Cristina A. Mendoza-Rodríguez, Michel Grutter, Thomas Blumenstock, and Frank Hase. Formaldehyde total column densities over mexico city: comparison between MAX-DOAS and solar absorption FTIR measurements. September 2020. doi: 10.5194/amt-2020-208. URL https://doi.org/10.5194/amt-2020-208.

[19] T. Carlund, N. Kouremeti, S. Kazadzis, and J. Gröbner. Aerosol optical depth determination in the UV using a four-channel precision filter radiometer. *Atmospheric Measurement Techniques*, 10(3):905–923, 2017. doi: 10.5194/amt-10-905-2017. URL https://www.atmos-meas-tech.net/10/905/2017/.

[20] A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2018005: Final Pandonia report, 2018.

[21] A. Cede and M. Tiefengraber. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2014007: Pandora-2S Maintenance - Quality Assurance Plan, 2014.

[22] A. Cede, M. Tiefengraber, M. Müller, M. Gebetsberger, D. Santana, A. Kreuter, and C. Posch. Workshop Report and Recommendations, 2019.

[23] L. Egli, J. Gröbner, G. Hülsen, L. Bachmann, M. Blumthaler, J. Dubard, M. Khazova, R. Kift, K. Hoogendijk, A. Serrano, A. Smedley, and J.-M. Vilaplana. Quality assessment of solar UV irradiance measured with array spectroradiometers. *Atmospheric Measurement Techniques*, 9(4):1553–1567, 2016. doi: 10.5194/amt-9-1553-2016. URL https://www.atmos-meas-tech.net/9/1553/2016/.

[24] B. Franco, F. Hendrick, M. Van Roozendael, J.-F. Müller, T. Stavrakou, E. A. Marais, B. Bovy, W. Bader, C. Fayt, C. Hermans, B. Lejeune, G. Pinardi, C. Servais, and E. Mahieu. Retrievals of formaldehyde from ground-based ftir and max-doas observations at the jungfraujoch station and comparisons with geos-chem and images model simulations. *Atmospheric Measurement Techniques*, 8(4):1733–1756, 2015. doi: 10.5194/amt-8-1733-2015. URL https://amt.copernicus.org/articles/8/1733/2015/.

[25] Manuel Gebetsberger, Jakob W. Messner, Georg J. Mayr, and Achim Zeileis. Estimation methods for nonhomogeneous regression models: Minimum continuous ranked probability score versus maximum likelihood. *Monthly Weather Review*, 146(12):4323 – 4338, 01 Dec. 2018. doi: 10.1175/MWR-D-17-0364.1. URL https://journals.ametsoc.org/view/journals/mwre/146/12/mwr-d-17-0364.1.xml.

[26] Andreas Geyer, Björn Alicke, Stephan Konrad, Thomas Schmitz, Jochen Stutz, and Ulrich Platt. Chemistry and oxidation capacity of the nitrate radical in the continental boundary layer near berlin. *Journal of Geophysical Research: Atmospheres*, 106(D8):8013–8025, 2001.

[27] V. Gorshelev, A. Serdyuchenko, M. Weber, W. Chehade, and J. P. Burrows. High spectral resolution ozone absorption cross-sections; Part 1: Measurements, data analysis and comparison with previous measurements around 293 K. *Atmospheric Measurement Techniques*, 7(2):609–624, 2014. doi: 10.5194/amt-7-609-2014. URL https://www.atmos-meas-tech.net/7/609/2014/.

[28] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[29] J. Herman, A. Cede, E. Spinei, G. Mount, M. Tzortziou, and N. Abuhassan. NO_2 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.

[30] G. Hönninger and U. Platt. Observations of bro and its vertical distribution during surface ozone depletion at alert. *Atmospheric Environment*, 36 (15):2481 – 2489, 2002. ISSN 1352-2310. doi: https://doi.org/10.1016/S1352-2310(02)00104-8. URL http://www.sciencedirect.com/science/article/pii/S1352231002001048. Air/Snow/Ice Interactions in the Arctic: Results from ALERT 2000 and SUMMIT 2000.

[31] J. Kaiser, G. M. Wolfe, K. E. Min, S. S. Brown, C. C. Miller, D. J. Jacob, J. A. deGouw, M. Graus, T. F. Hanisco, J. Holloway, J. Peischl, I. B. Pollack, T. B. Ryerson, C. Warneke, R. A. Washenfelder, and F. N. Keutsch. Reassessing the ratio of glyoxal to formaldehyde as an indicator of hydrocarbon precursor speciation. *Atmospheric Chemistry and Physics*, 15(13): 7571–7583, 2015. doi: 10.5194/acp-15-7571-2015. URL https://acp.copernicus.org/articles/15/7571/2015/.

[32] A. Kreuter. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2018003: Recommendations on operational AOD measurements for Pandonia, 2018.

[33] J. Lampel, Y. Wang, A. Hilboll, S. Beirle, H. Sihler, J. Pukite, U. Platt, and T. Wagner. The tilt effect in DOAS observations. *Atmospheric Measurement Techniques*, 10(12):4819–4831, 2017. doi: 10.5194/amt-10-4819-2017. URL https://www.atmos-meas-tech.net/10/4819/2017/.

[34] King-Fai Li, Ryan Khoury, Thomas J Pongetti, Stanley P Sander, Franklin P Mills, and Yuk L Yung. Diurnal variability of stratospheric column no 2 measured using direct solar and lunar spectra over table mountain, california (34.38° n). *Atmospheric Measurement Techniques*, 14(12):7495–7510, 2021.

[35] Kiyoshi Matsumoto and Hiroshi Tanaka. Formation and dissociation of atmospheric particulate nitrate and chloride: an approach based on phase equilibrium. *Atmospheric Environment*, 30(4):639–648, 1996.

[36] Brendan McAndrew, Joel McCorkel, Timothy Shuman, Barbara Zukowski, Aboubakar Traore, Michael Rodriguez, Steven W. Brown, and John T. Woodward. Goddard Laser for Absolute Measurement of Radiance for Instrument Calibration in the Ultraviolet to Short Wave Infrared. *2018 Conference on Lasers and Electro-Optics (CLEO)*, pages 1–2, 2018.

[37] M. Müller, M. Tiefengraber, and A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2016011: Validation reports, 2016.

[38] Daniel Pérez-Ramírez, David N. Whiteman, Alexander Smirnov, Hassan Lyamani, Brent N. Holben, Rachel Pinker, Marcos Andrade, and Lucas Alados-Arboledas. Evaluation of AERONET precipitable water vapor versus microwave radiometry, GPS, and radiosondes at ARM sites. *Journal of Geophysical Research: Atmospheres*, 119(15):9596–9613, August 2014. doi: 10.1002/2014jd021730. URL https://doi.org/10.1002/2014jd021730.

[39] A. J. M. Piters et al. The Cabauw Intercomparison campaign for Nitrogen Dioxide measuring Instruments (CINDI): design, execution, and early results. *Atmospheric Measurement Techniques*, 5:457–485, 2012.

[40] Armistead G Russell, Glen R Cass, and John H Seinfeld. On some aspects of nighttime atmospheric chemistry. *Environmental science & technology*, 20 (11):1167–1172, 1986.

[41] Gregory L. Schuster, Oleg Dubovik, and Brent N. Holben. Angstrom exponent and bimodal aerosol size distributions. *Journal of Geophysical Research: Atmospheres*, 111(D7), 2006. doi: 10.1029/2005JD006328. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006328.

[42] H. Slaper, H. A. J. M. Reinen, M. Blumthaler, M. Huber, and F. Kuik. Comparing ground-level spectrally resolved solar UV measurements using various instruments: A technique resolving effects of wavelength shift and slit width. *Geophysical Research Letters*, 22(20):2721–2724, 1995. doi: 10.1029/95GL02824. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95GL02824.

[43] E. Spinei, A. Whitehill, A. Fried, M. Tiefengraber, T. N. Knepp, S. Herndon, J. R. Herman, M. Müller, N. Abuhassan, A. Cede, P. Weibring, D. Richter, J. Walega, J. Crawford, J. Szykman, L. Valin, D. J. Williams, R. Long, R. J. Swap, Y. Lee, N. Nowak, and B. Poche. The first evaluation of formaldehyde column observations by pandora spectrometers during the korus-aq field study. *Atmospheric Measurement Techniques Dis-*

cussions, 2018:1–35, 2018. doi: 10.5194/amt-2018-57. URL https://www.atmos-meas-tech-discuss.net/amt-2018-57/.

[44] 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 Measuerement Techniques Discussions, in review*, 2020.

[45] R. Spurr, D. Loyola, M. van Roozendael, C. Lerot, K.P. Heue, and J. Xu. S5P/TROPOMI Total Ozone Algorithm Theoretical Baseline Document. S5P-L2-DLR-ATBD-400A, 2018. URL http://www.tropomi.eu/sites/default/files/files/Sentinel-5P-TROPOMI-ATBD-Total-Ozone.pdf.

[46] S. Studer, K. Hocke, A. Schanz, H. Schmidt, and N. Kämpfer. A climatology of the diurnal variations in stratospheric and mesospheric ozone over bern, switzerland. *Atmospheric Chemistry and Physics*, 14(12):5905–5919, 2014. doi: 10.5194/acp-14-5905-2014. URL https://www.atmos-chem-phys.net/14/5905/2014/.

[47] 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.

[48] 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.

[49] M. Tiefengraber and A. Cede. ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study, LuftBlick Report 2017002: Network intercomparison campaign report, 2017.

[50] J. Timmer and M. Koenig. On generating power law noise. *aap*, 300:707, Aug 1995.

[51] 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 Roozendael, 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.

[52] Y. Wang, A. Apituley, A. Bais, S. Beirle, N. Benavent, A. Borovski, I. Bruchkouski, K. L. Chan, S. Donner, T. Drosoglou, H. Finkenzeller, M. M. Friedrich, U. Frieß, D. Garcia-Nieto, L. Gómez-Martín, F. Hendrick, A. Hilboll, J. Jin, P. Johnston, T. K. Koenig, K. Kreher, V. Kumar, A. Kyuberis, J. Lampel, C. Liu, H. Liu, J. Ma, O. L. Polyansky, O. Postylyakov, R. Querel, A. Saiz-Lopez, S. Schmitt, X. Tian, J.-L. Tirpitz, M. Van Roozendael, R. Volkamer, Z. Wang, P. Xie, C. Xing, J. Xu, M. Yela, C. Zhang, and T. Wagner. Inter-comparison of max-doas measurements of tropospheric hono slant column densities and vertical profiles during the cindi-2 campaign. *Atmospheric Measurement Techniques*, 13(9):5087–5116, 2020. doi: 10.5194/amt-13-5087-2020. URL https://amt.copernicus.org/articles/13/5087/2020/.

[53] Mark Wenig, Bernd Jähne, and Ulrich Platt. Operator representation as a new differential optical absorption spectroscopy formalism. *Appl. Opt.*, 44(16):3246–3253, Jun 2005. doi: 10.1364/AO.44.003246. URL http://ao.osa.org/abstract.cfm?URI=ao-44-16-3246.

[54] S. N. Wood. Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, 65(1):95–114, 2003.

[55] S. N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36, 2011.

[56] S.N Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2 edition, 2017.

[57] 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/.

[58] Y. Zong, S. W. Brown, B. C. Johnson, K. R. Lykke, and Y. Ohno. Simple spectral stray light correction method for array spectroradiometers. *Applied Optics*, 45(6):1111–1119, 2006.

2 Processing software evolution

The Pandora spectrometer system (Pandora) was initially designed as a campaign instrument to support satellite validation and air quality studies. The first operational software of the Pandora was the *Pan Software Suite* (PSS). The evolution of the instrument to a monitoring instrument (summarized in the final Pandonia report [20]) made it necessary to implement a number of new hardware and software features in order to meet the demands on data quality, availability and traceability. The new software features are incorporated in an evolution of the PSS, the *Blick Software Suite* (BSS). At the start of this project in August 2018 the BSS was already the operational software. When we refer to the BSS in the text, we refer to version 1.8.

In Table 1 some key aspects between PSS and BSS regarding data handling are compared.

3 Processing algorithms and data products

3.1 Development tool: SIMPLE

When new algorithms are designed or existing ones optimized, our development tool SIMPLE (SIMulate Pandora data LEvels) comes into use. With SIMPLE we can test e.g. the impact of different fitting parameters such as the selected wavelength range, polynomial orders, stray light correction method and many others on the error of total column amounts using synthetic data. Hence SIMPLE is able to describe the "algorithm error", i.e. what the systematic error in the data products will be even if no instrumental uncertainties other than noise, which can be included in SIMPLE, are present. Within the FRM4AQ project, SIMPLE is an important tool to determine the best retrieval settings for a data product.

3.2 Calibration improvements

This section describes new L1 data correction steps which benefit any new algorithms.

3.2.1 Absolute calibration

The correction of the instrument sensitivity (absolute calibration) for all filter combinations is implemented in the BSS. Retrieval algorithms utilizing a literature reference spectrum in the spectral fitting benefit in particular. In these cases low order closure polynomials describe solely atmospheric conditions and therefor improve homogeneity of certain data products (as e.g. shown for total O3 in *Tiefengraber and Cede* [48]). High quality absolute calibration is also a basic requirement for AOD retrieval.

3.2.2 Matrix stray light correction

A spectral stray light (SpecSL) correction method similar to the so-called *matrix* SpecSL method [e.g. 58] has been included in the BSS. The correction is implemented in a fully analytical way using parameterized line spread functions (LSF) extracted from laser measurements. To this end the full slit function can be represented by 4 sets of parameters: core parameters, stray light parameters, transition

Table 1: Comparison of data handling between the different software releases.

	PSS	BSSv1.7	BSSv1.8
		DATA LOGISTICS	
Policy	Generic disclaimer (principal investigator not known)	Site application form	same as BSSv1.7
Processing frequency	Daily	Real time (≈ 10 min.)	same as BSSv1.7
Station naming	Any name allowed	"Intuitive read" criteria (e.g. IBK ⇒ Innsbruck)	same as BSSv1.7
		TRACEABILITY	
Calibration validity	Not implemented (manual renaming)	Validity period and version fully implemented for processing	same as BSSv1.7
Data files naming	Not uniform (dependent on calibration file)	Uniform description with versioning and provided/updated by PGN	Calibration validity period added to file name.
		QUALITY CONTROL, QUALITY CHECK	
Raw data	Not implemented	First check upon data arrival at server	Front-end diagnosis plots added
Retrieval products	Not implemented	Each data level categorized in high, medium, low quality data	Automated daily statistical break point analysis added
		DATA	
Instrument characterization	Common characterization steps (see e.g. *Cede and Tiefengraber* [21])	major additions: matrix stray light correction, absolute calibration, micro window wavelength correction for reference spectra.	major additions: stray light correction includes structures (e.g. "ghost"), improved wavelength correction accuracy (up to 1 pm)
Gas reference amount calibration	MLE	MLE, E-MLE	MLE, E-MLE, AXC
Uncertainty treatment	propagated measurement noise	same as PSS	uncertainty categories added: independent, structured and common uncertainty. E.g. calibration uncertainty or impact of wrong cross section temperature considered.
Climatologies, databases	none	Rayleigh AMF lookup-tables (for sky)	O3 profile climatology (TOMS v8 [16]), NO_2 stratospheric column climatology (OSIRIS [17]), temperature/pressure climatology (COSPAR [13])
Product categories	Total columns from sun	Total columns from sun and moon, tropospheric columns from sky data	Total columns from sun and moon, tropospheric columns and profiles from sky data

parameters (defining transition between core and stray light region) and stray light special region parameters (find details in *Cede* [6]). The stray light matrix for all pixels (wavelengths) of e.g. Pandora 31s1, with one example LSF at 450 nm, can be seen in Figure 1. The paramterizations make it also possible to account for the pronounced ("ghost") feature about 120 nm left of the exciting wavelengths and allow a varying baseline value (note the bluish - purplish variation in the background).

Despite very careful laboratory measurements, we initially noticed that this new SpecSL correction led to inconsistencies in some data products. In depth investigations revealed that some presumably high quality diode lasers turned out to have a drift of the output wavelength over a short time (several seconds), which causes a too wide measured resolution, which introduces an error when fitting the LSFs. When including those faulty lasers in the analysis, the spectral inter- and extrapolation of the LSFs over the entire wavelength range of the instrument was erroneous. A comparison of the diode lasers used at the Pandora lab at GSFC to a reference tunable laser system GLAMR [36] helped to reveal this issue. This is shown in figure 2, where the difference between the retrieved resolutions from the laser are compared to the ones determined from spectral lamps, which is the standard procedure. As can be observed, the difference for the GLAMR laser (bottom panel)

is very small, but for some of the diode lasers (375, 405 and 445 nm) it is very pronounced.

Figure 1: Parameterized stray light matrix as characterized for Pandora 31s1. The inhomogeneity of the stray light distribution is evident and justifies the inclusion of parameterizing also "special regions". The central figure shows all line spread functions (for the respond wavelengths in the x-axis) triggered by all exciting wavelengths (y-axis). One explicit example of a line spread function is extracted for the exciting wavelengths at 450 nm in the top figure. Here the so-called "ghost" special region is clearly visible between 300 and 350 nm. Note, if one would plot all readings in y-direction (along the exciting direction) for a certain respond wavelength, this function would be called filter function. So to say the collection of intensities "landing" at the respond wavelength.

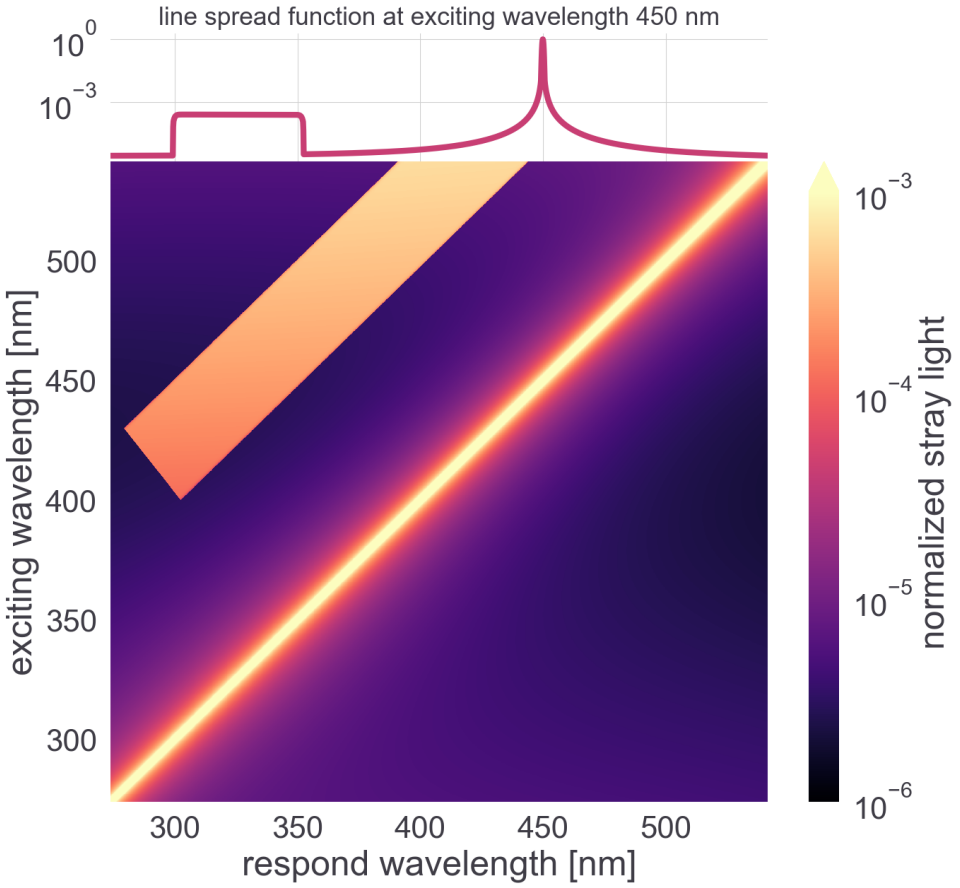
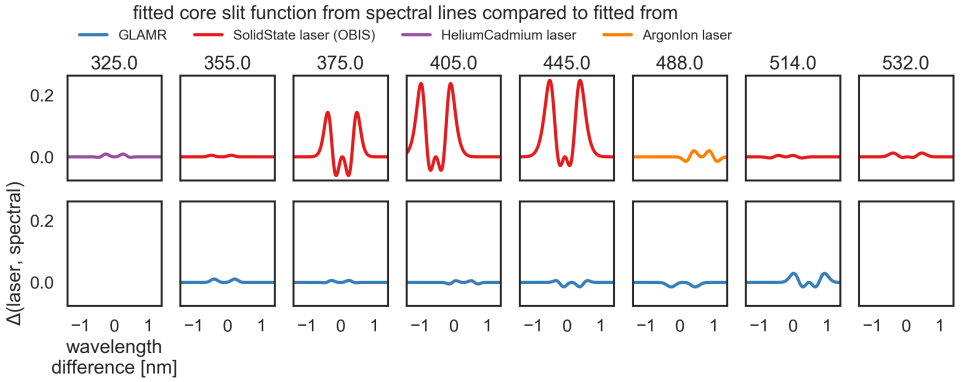
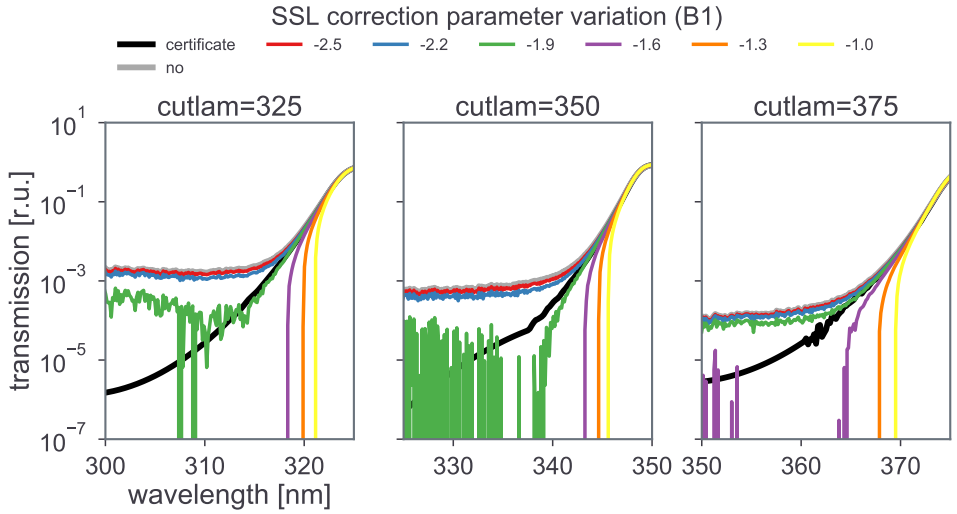


Figure 2: Difference of slit functions retrieved from spectral lines and different lasers types: solid state lasers (red), helium cadmium laser (purple), argon ion laser (orange) and the GLAMR laser (blue). The solid state lasers at 375, 405 and 445 nm are broadened and deviate from the spectral lines.



Besides possible technical issues with lasers, laser measurements themselves are very critical to perform. First the laser setup is very important and in order to have a good signal-to-noise in the regions away from the peak wavelength, laser lines are also measured with saturation. All this adds to possible difficulties and uncertainties for the retrieval of the slit function parameters from laser measurements.

Figure 3: Transmission measurements for cut-on filters at 325, 350 and 375 nm (from left to right): The black line represents the certificate (reference), determined with a double monochromator. Color-coded are transmissions calculated with variations of the strongest SpecSL parameter (B1). Comparison ot the certificate can serve as SpecSL parameter calibration. For comparison, the transmission calculated without SpecSL correction is shown in gray.



This is why we tested a different approach to determine a SpecSL correction. By utilizing a combination of band-pass and cut-on filters at measurements with an FEL lamp, we want to retrieve the SpecSL parameters in another way. Figure 3 illustrates the principle. Shown are transmission measurements for three different cut-on filters (325, 350, 375 nm), i.e. the ratio between the light output without with the filter in place over the ones without the filter in place. Then the SpecSL parameters are determined by varying them, applying the stray light correction on the measurements (gray), ad comparing the result to the "ideal" signal as given by the certificate (black), which gives the cut-on transmission determined by the manufacturer with a stray light free double monochromator system. The colored lines in figure 3 show the results for different values of the most important SpecSL parameter B1. It shows that for a value of around -1.9 for B1 the corrected measurements match the certificate best.

3.2.3 L1 wavelength displacement determination based on micro windows

The technique to determine the "true" wavelength grid of a spectrum implemented in the PSS is based on a comparison to a solar atlas. In May 2019, the BSSv1.7 was extended by another method based on spectral fitting, which utilizes micro windows across the entire spectral region. In this way trace gas absorption features are also taken into account in the wavelength change retrieval.

3.2.4 L1 wavelength displacement determination based on long windows

Starting with BSSv1.8, the wavelength displacement was further improved and is now estimated based on spectral fitting across the hole spectral range (one long fitting window), considering all major absorber. The output of the spectral fitting can be utilized, because in BSSv1.8 all “tilt-effects” (that is algorithm errors, details in section 3.5.1) are considered.

3.3 Direct sun total NO_2

The current operational total NO_2 is already at a very high quality level. Precision-wise total NO_2 performs exceptionally good (better than about 0.5 %, compare e.g. *Tiefengraber and Cede* [47], *Zhao et al.* [57]). The current way of calibrating NO_2 [1] is the major driver of an estimated accuracy of 0.1 DU.

The soft-calibration technique used at present is the *Minimum Langley Extrapolation* (MLE, details e.g. *Herman et al.* [29]). The main assumption of the MLE is that a certain subset of the data used for calibration only contains a background value of the gas (e.g. only stratospheric NO_2), which further is assumed to be constant over the day. Sub-setting the data to meet these criteria[2] is the major uncertainty source in the MLE and is directly connected to the variability of the tropospheric gas content during the calibration period.

[1] A measured spectrum is selected as reference. For this reference spectrum the NO_2 content needs to be found.

[2] Subset the data in air mass factor bins and define for each bin a certain percentile to be representative for the background value

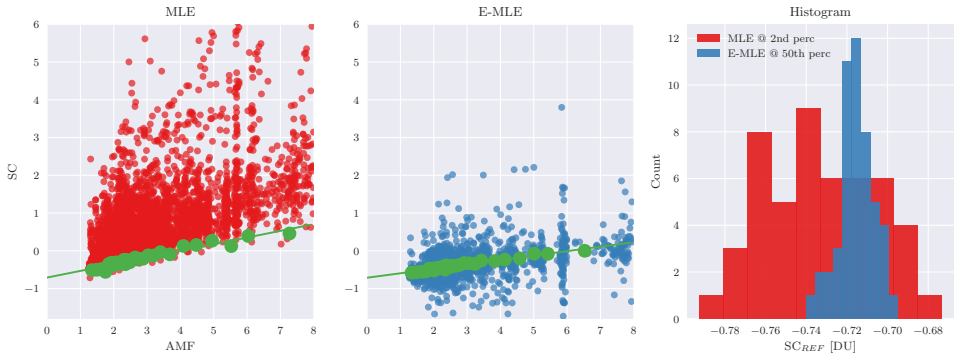
In BSSv1.8 we have included an *extended* MLE technique (E-MLE) where the (variable) tropospheric content taken from sky observations (see section 3.14) is removed from the total columns during the calibration process. By this, per definition, all data points are converted to valid MLE candidates. This does not only significantly remove the MLE uncertainty (simplified data sub-setting), but also very short datasets can be used to perform a calibration (e.g. one clear sky week instead of one clear sky month).

First tests have been applied to Pandora 117 (deployed in Rome) and are shown in Figure 4. In the left figure panel the MLE analysis plot is drawn with the retrieved relative NO_2 slant columns in red, the selected minimum values (usually assumed to be found at the 2nd percentile) of each air mass factor (AMF) bin (here 30 bins) in green, with the linear fit extrapolated to zero AMF (equals the SC in the reference). The middle figure panel shows the same but for the E-MLE, that is each relative slant column is a stratospheric slant column now. Consequently the linear fit is applied to the median value of each AMF bin, since all data sets fulfill the criteria (constant background).

The selection of the number of bins particularly impacts the linear fit, as it is shown in the right figure panel by the width of the distributions (for the MLE the common 2nd percentile scenario is shown). At least for this example, the added uncertainty by selecting the number of bins is reduced to about one third compared to the regular MLE.

The E-MLE relies on high quality tropospheric NO_2 columns, which are provided by the BSS (see section 3.14). The MLE has also been tested at stations with different NO_2 conditions (e.g. urban, suburban, remote) and has proven to work well. Therefore is has been implemented in BSSv1.8.

Figure 4: Total NO_2 calibration performance for MLE (left panel) and E-MLE (middle panel) is shown (explanation in text). The green dots emphasize the values where the MLE criteria are met: the 2nd (50th) percentile value in each AMF bin for the MLE (E-MLE). The slant column in the reference for different numbers of AMF bins is shown in the right panel for MLE (in red) and E-MLE (in blue).



3.3.1 Direct sun total NO_2 from spectrometer 2

Within the PGN there are Pandora-1S instruments and 2S instruments. 1S instruments cover the wavelength range from 270 nm to 540 nm and 2S instruments go from 270 nm to 900 nm by having two different spectrometers (S1 from 270 nm-540 nm, S2 from 390 nm-940 nm) with two individual optical paths with both spectrometers covering the NO_2 retrieval region from 400 nm to 470 nm. Obviously, one motivation to present NO_2 from the S2 is to have additional measurements: for typical 2S instruments, if the S1 measures with the diffuser on filterwheel 2 position 2 (called DIFF2 in the following), the S2 measures with the diffuser on filterwheel 2 position 5 (called DIFF5 in the following) and if the S1 measures with the U340 bandpass filter, S2 measures with DIFF2. From both diffuser measurements of S2, NO_2 can be retrieved. These additional measurements can be used to fill gaps which are currently present when just using S1 measurements. Further, the S2 measurements can be used to identifiy potential drifts in one of the spectrometers. Another motivation to work with S2 is to identify shortcomings in the software, hardware and calibration, also with respect the goal to retrieve other gases from S2 in the future (e.g. H20). The used retrieval codes for S2 are the same as for S1.
Two Pandora 2S-instruments have been co-located in Rome-SAP from end of July 2020 till mid October 2020. This is a good opportunity to compare two instruments with the focus on the S2.

From first data comparisons between DIFF2 and DIFF5 measurements we already know that the data of the measurements of the diffuser, which has not been used for the reference, are biased and also the root mean square of the spectral fitting residuals (RMS) increased. This means, if measurements of DIFF2 are taken as reference measurements, the DIFF5 measurements have a significantly higher RMS and are biased regarding the total column NO_2. The reason for these observations are shortcomings in the L1 calibration, more precisely in the sensitivity calibration, where spectral features in the diffusers cannot be characterized sufficiently. We could not yet identify the reason for this issue, but we reckon it to be connected to the interplay between input beam divergence (sun vs. FEL) and diffusers.

The best opportunity to directly compare the S2 NO_2 data with S1 NO_2 data are the DIFF5 measurements of the S2, because of the hardware setup these are the measurements which are taken simultaneously. Due to a higher sensitivity of the S2 and the internal ND filter combination, the integration time for a measurement differs between S1 and S2 and consequently the amount of measurements between S1 DIFF2 and S2 DIFF5 are different and the measurements have been averaged in time. In order to be independent of MLE calibration uncertainties (see 3.3), uncalibrated NO_2 SCs are used for the comparison. The results are shown in figure 5. Further results how the datasets agree towards P138s1 are shown in table 2. The correlation of 1 for both cases is excellent, the offset is small and in the same order of magnitude for both cases and the slope for Pandora 117 is excellent but a little higher for Pandora 138. In this example, DIFF5 measurements have been used for the reference measurement. This clearly shows that the S2 measurements of at least one diffuser can be used as additional NO_2 measurements, complementing the measurements of S1.

Since the absolute S2 NO_2 values will be determined by calibrating using the S1 as reference, eliminating the MLE as main driver for the uncertainty, it can be estimated to have the same accuracy as for S1. As outlined in *Tiefengraber and Cede* [47], the drivers for the precision are USS, SL and noise. Due to the same optical setup of the S2, USS can be assumed to be the same and SL does not play a critical role so far regarding NO_2. Due to recent upgrades in the headsensor design, using finer ND filter steps for new instruments, the noise can be expected to be lower for the S2. Consequently, the precision of the S2 is expected to be slightly better as for S1.

Figures 6 and 7 show how the S2 DIFF2 and DIFF5 measurements are biased towards each other and how the S2 measurements can complete the S1 measurements.

Conclusions

- We suggest that the S2 measurements of just one diffuser are used for the moment, most useful would be the diffuser which measures when the S1 is doing U340 measurements. The measurements of the other diffuser should be flagged.
- We got an idea about shortcomings in the L1 correction, which would be interesting to solve with regard to other trace gases and especially regarding AOD.

Table 2: Comparison table of relative slant columns for four different spectrometers towards Pandora 138s1 as reference dataset. The S2 data have been split up into both diffuser measurements which are used in the field. Data have been averaged by 10min.

P138s1	**slope**	**intercept** $[mmol/m^2]$	**CC**	**N**
P138s2 DIFF2	0.999	4e-3	0.99	1630
P138s2 DIFF5	1.029	1.8e-2	1	2112
P117s1	0.981	-1e-3	0.994	1771
P117s2 DIFF2	0.979	-2.8e-2	0.993	1783
P117s2 DIFF5	1.002	-1.2e-2	0.994	1862

Figure 5: 2min averaged measurements from 1st August till 15th September of Pandoras 117 and 138. The intercept is y, r is the correlation coefficient and N shows the number of filtered measurements (just quality flag 0 data allowed) for both spectrometers.

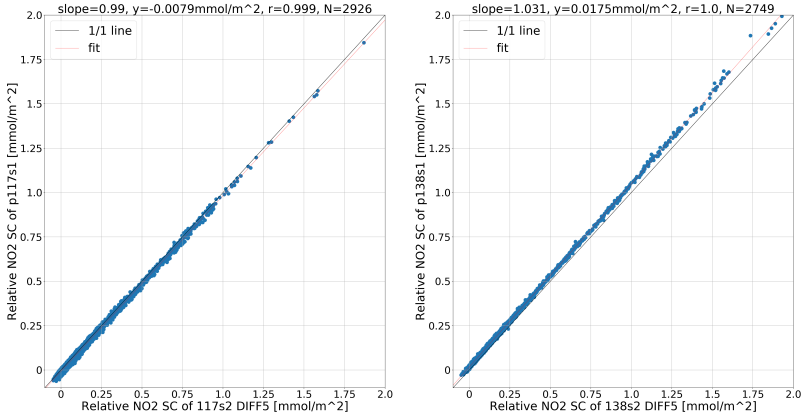


Figure 6: Data shown for P117 for both spectrometers, where the instrument was set on a dedicated high resolution measurement schedule. The DIFF2 measurements can fill the gaps when the S1 measurements are done with the U340 filter.

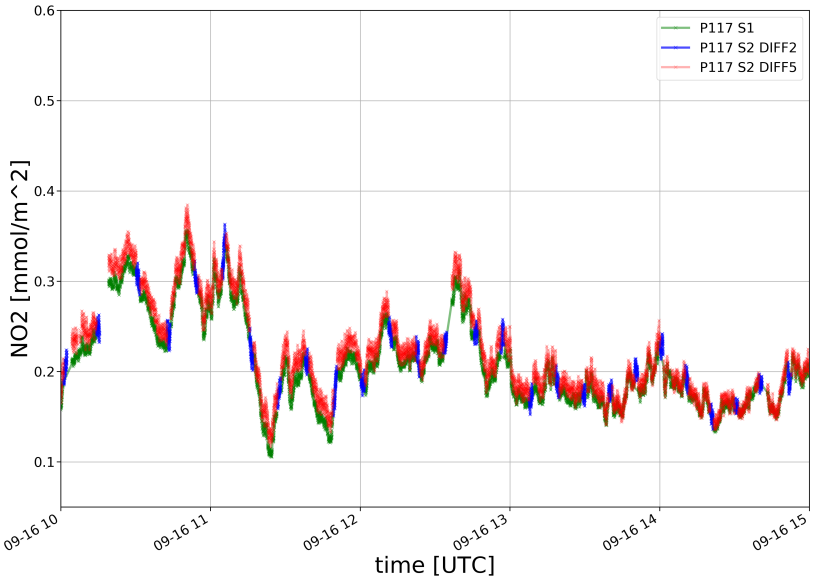
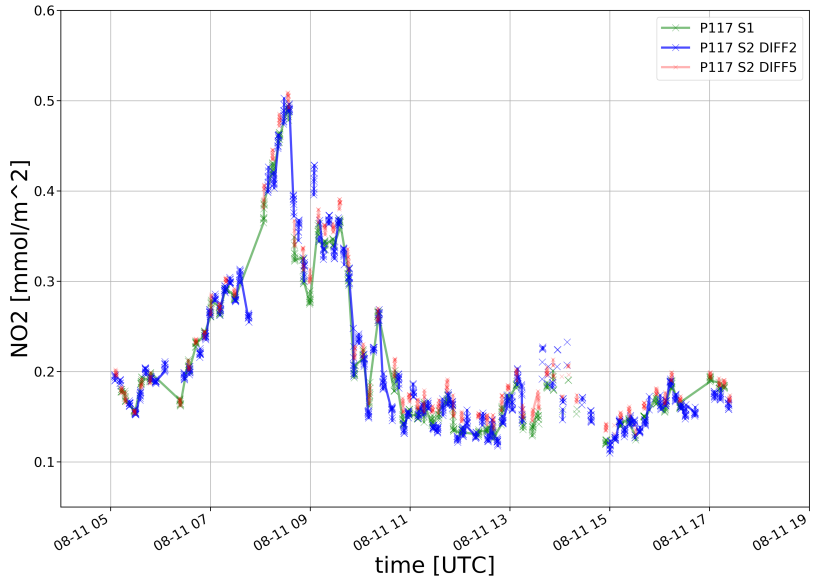


Figure 7: Data shown for P117 for both spectrometers for one day on a regular schedule. S1 measurements and S2 DIFF5 measurements are taken at the same time and the DIFF2 measurements are in between. The S2 DIFF5 and S2 DIFF2 measurements are biased.



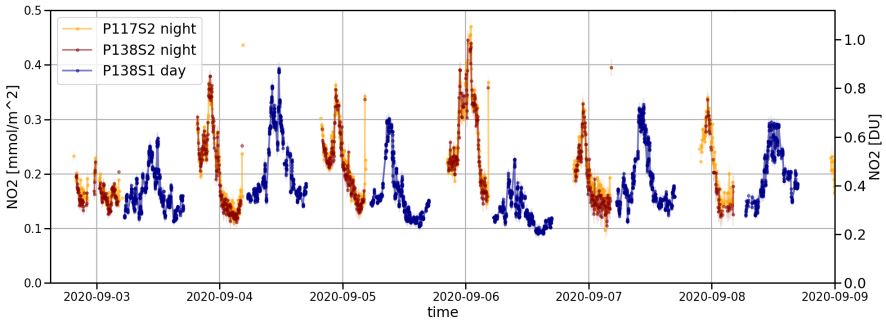
3.4 Direct moon total NO_2

Direct moon (DM) data is retrieved between 400 nm and 470 nm, with a 4th order background polynomial and a stratospheric-tropospheric split applied as for the direct sun retrieval. This setup has shown the best comparison results between collocated instruments and is insensitive against sudden changes in the hardware, e.g. slight alignment problems.

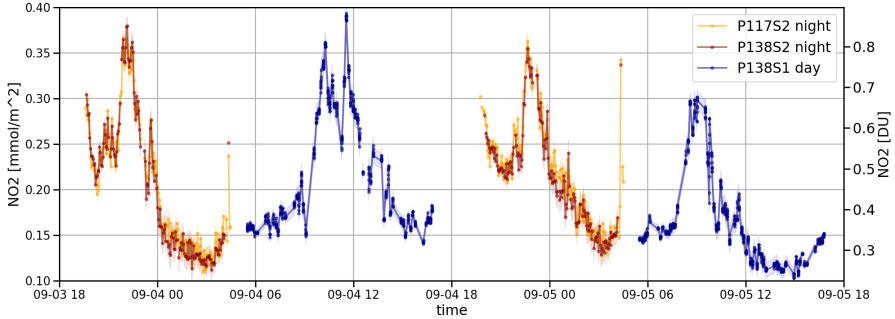
Moon data of both Pandora spectrometers (S1=UV, S2=VIS) can be retrieved with a direct sun reference spectrum and the slant column value is determined by a MLE with direct sun data. Hence, the field calibration of the direct sun NO_2 is being used and the moon data can be processed without additional work. Further, the uncertainty from the calibration (common uncertainty) is the same as for direct sun data. Since spectrometer 2 is calibrated using the spectrometer 1 as reference, it has the same calibration uncertainty, with an additional uncertainty coming from the calibration towards spectrometer 1 which is expected to be small. The instrumental uncertainty however (noise) is lower for the spectrometer 2 due to the higher sensitivity.

As further shown in section 4.2 and as can be seen in figure 8, the lunar TotNO_2 data agrees well within the uncertainty and can resolve highly variable NO_2 pollution events. If available, it is recommended to use the spectrometer 2 direct moon data, but spectrometer 1 data can also be used with the restrictions due to the lower sensitivity, i.e. higher uncertainty and fewer datapoints.

Figure 8: Lunar and direct sun TotNO_2 of Pandora 117 and 138. The original direct sun data with quality flag 0 is shown and lunar data is averaged by 3min with just using data with a total uncertainty below 0.02mmol / m^2 and a wrms limit of 2e-3. Due to the lower noise, spectrometer 2 is shown for lunar data which agrees well within the uncertainty.



(a) Lunar and direct sun TotNO_2 for a few days in Rome-SAP.



(b) Close up for the same dataset as above.

3.5 Direct sun total O_3 and O_3 temperature

The current operational total O_3 product is in general not very accurate due to the following reasons:

- The nominal fitting window is set to 310 to 330 nm. This window was never assessed in a sophisticated analysis as e.g. shown in section 3.1 and might be not the best wavelength range to be used.
- The retrieval is based on laboratory calibration only. This means the measured spectra are referenced to a literature spectrum ("out of the lab" ozone).
- Absolute calibration is not applied for most units. Due to the use of a literature reference (in the retrieval), any non-smooth structure in the instrument's spectral sensitivity inside the O_3 wavelength range in general causes a bias in the data.
- Elaborate SpecSL calibration is not applied to most instruments, which gives a negative bias in total O_3 starting at solar zenith angle (SZA)=70 deg and sometimes even at smaller values.
- We assume a fixed effective O_3 temperature (O3temp) of 225 K in the algorithm, which commonly leads to underestimation (overestimation) of total O_3 in summer (winter) [57]. In addition to this bias in total O_3, incorrect O_3temp assumptions also bias SO_2 retrievals.

Above mentioned points are discussed in more detail e.g. in *Müller et al.* [37] and *Tiefengraber and Cede* [48].

A major step to improve total O_3 is the incorporation of O_3temp in the spectral fitting. The basic requirement in order to do this is to use a measured reference spectrum instead of the literature reference, since the temperature dependence of the O_3 cross sections is rather weak and can easily by overshadowed by the differences between the Pandora spectral sensitivity and the literature reference spectrum. Note that the main challenge to retrieve absolute values of total O_3 and O_3temp is to determine the total O_3 column and O_3temp in the chosen reference itself. Once this is known, the retrieval is rather straight forward.

3.5.1 Direct sun total O_3temp

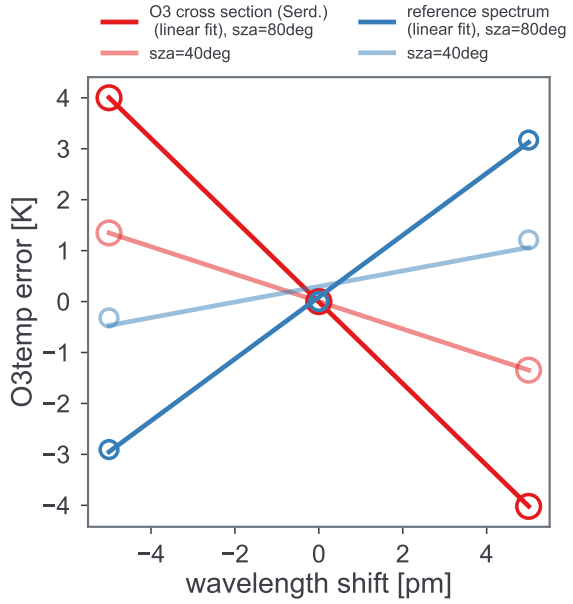
A bias of 1 K in O_3temp translates to about 1 DU SC error and hence is significant. The annual variation of O_3temp increases with latitude[3] which makes seasonal biases in O_3 columns a significant problem due to the fact that still the majority of instrument sites are located in the mid latitudes.

On the other hand one can expect the daily variation of O_3temp to be insignificant in our columnar perspective, mainly for the reason that O_3temp is strongly linked to the vertical profile of O_3 which emphasizes stratospheric O_3 abundances. In addition, the diurnal variation of O_3 concentrations in the upper stratosphere (mesosphere) is inverse to the lower stratosphere, which also has a compensating effect (this is linked to the availability of O radicals) [46].

Sensitivity tests performed with SIMPLE and also with real measurements revealed a rather strong impact of the exact wavelength registration of the reference spectrum on the retrieved O_3temp, which was not expected by us. Even a registration error in the order of some pico meters (pm) already significantly biases O_3temp, as illustrated in Figure 9 (blue lines). For this simulation, the reference is compiled from a noon spectrum and displaced from -5 to +5 pm (note that also a wavelength change polynomial is included in the spectral fitting). The impact is shown in red (at 40 deg and 80 deg solar zenith angles).

[3] Based on TOMS climatology of O_3temp, annual peak-to-peak variation range from about 5 K at 20 deg latitude to about 10 K for mid latitudes and even up to 25 K for high latitudes.

Figure 9: Simulated O_3temp errors as a function of spectral shifts (from -5 to +5 pm) in a synthetic reference spectrum (blues) and the Serdyuchenko O_3 cross sections (reds). This is shown for SZA 80 deg (bold colors) and 40 deg (faint colors) to emphasize the AMF dependency. Solid lines are linear fits in the data points.



Besides that, we followed a hint brought up by Michel van Roozendael (BIRA-IASP) at the 1st PGN workshop in Innsbruck, Austria [22] (and personal communications with Richard Siddans (RAL Space)) that the Serdyuchenko O_3 cross sections [27] might be shifted by 3 pm[4]. Therefore we tested the magnitude of this effect ourselves, which is depicted in red (again at 40 deg and 80 deg) in the figure. We observe the strongest effect on O_3temp caused by the O_3 cross section shift for high solar zenith angles and similar but reversed impact for a shifted reference spectrum, where it is in the range of 0.5 to 1 K/pm. This is a very significant impact and would further translate to about 0.5 to 1 DU bias in total O_3 per 1 pm error in the wavelength registration.

[4]As stated in *Gorshelev et al.* [27], different spectrometers (Echelle and Fourier Transform technique) were used to determine the cross sections and for both a wavelength calibration accuracy of 5 pm is given. In their study they used the Echelle spectrometer below 310 nm and the Fourier Transform spectrometer between 310 and 350 nm.

This is a particular challenge, because wavelength registration accuracies in the order of 1 pm has to our knowledge never been considered significant in the DOAS or "irradiance" community and hence techniques for dispersion determination or its change are not at that level yet.

As a consequence we had to tackle two things:

1. Improve our technique to determine the correct wavelength registration for our spectra.

2. Investigate a possible wavelength displacement of the Serdyuchenko O_3 cross sections.

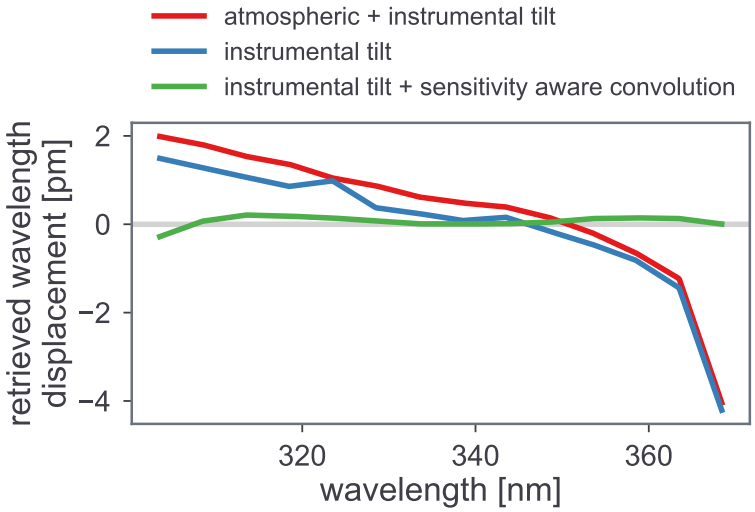
Wavelength displacement correction

The best possible wavelength registration of the spectra is of crucial importance. This is true in particular for the reference spectrum (as seen e.g. in Figure 9). In the DOAS community, wavelength shifts in the spectrum are usually captured by allowing a wavelength correction polynomial in the spectral fitting. And for the compilation of reference spectra, a running micro window approach is applied to get wavelength displacement information for the full spectrum (e.g. used by QDOAS). Besides the obvious difficulty to have appropriate fitting setups available for each micro window, an intrinsic bias in the estimated wavelength displacement remains: since broadband atmospheric transmission and instrumental features do not commute with the convolution of the instrument slit function [53], spectral features are generated. This effect is usually referred to as *tilt-effect* [33] and manifests as apparent wavelength shift in the order of several pm. Meaning, owing to the standard approach in the DOAS world of using pre-convoluted cross sections, the magnitude of the evaluated wavelength change polynomial does *not* represent the "true" wavelength shift of the spectra.

This effect is illustrated in Figure 10 for a simulated spectrum with a band pass filter (U340). Displayed is the outcome of the wavelength correction polynomial from the spectral fitting for a un-shifted spectrum. Although actually no wavelength displacement has happened, the algorithm reacts on the tilt introduced by atmospheric absorber and the instrument sensitivity (red line). The tilt is reduced if atmospheric absorption is removed (blue line), but the instrumental tilt is still dominating. If however the instrument sensitivity is already taken into account during convolution, the tilt effect vanishes (green line). This has been implemented in

BSSv1.8.

Figure 10: Apparent retrieved wavelength shift from spectral fitting due to the tilt-effect. Three cases are shown: Atmospheric + instrumental tilt (red), only instrumental tilt (blue) and instrumental tilt but with sensitivity aware convolution (green).

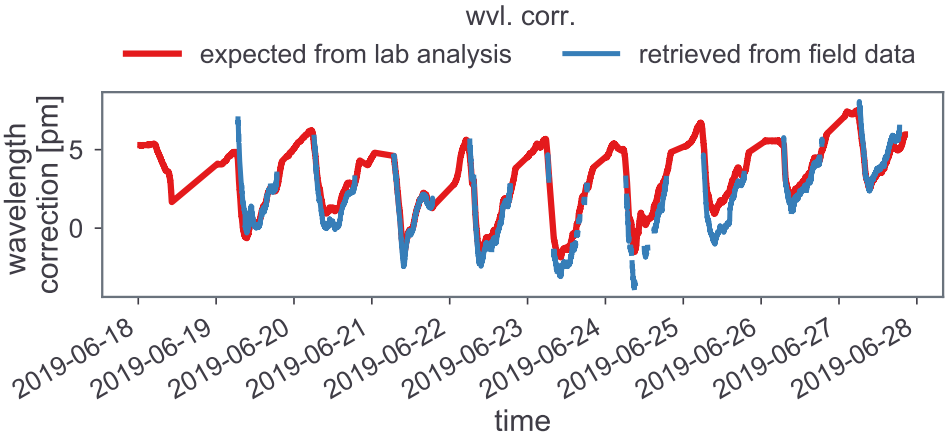


Also commonly used, preferably in the "irradiance" community, is the comparison of measured spectra to a solar atlas (this is also done by default for Pandora L1 data until BSS version 1.7) [e.g. 42, 23]). Since for this approach usually no gas absorption is taken into account biases typically arise in spectral regions with strong gas absorption, as in the UV due to O_3 absorption.

In order to apply all necessary corrections to prevent above explained errors in the wavelength displacement correction, the BSS has been extended quite substantially. The quality of the new wavelength correction approach is shown in Figure 11 for a Pandora 121 in Huelva, June 2019. The diurnal variation of the wavelength displacement, retrieved from field data by spectral fitting, is displayed in blue. In contrast, the red line in this figure depicts the wavelength displacement based on temperature variations only, as it was characterized in the laboratory [5]. Evidently, the wavelength correction boils down to compensate the dispersion change due to environmental temperature variations. This might sound obvious, is however a prove that the wavelength registration correction in the BSS was done properly. Please note the small scales of average diurnal amplitudes of about 5 pm.

[5]This characterization is done by measuring the spectral dispersion as retrieved from spectral lamps for different instrument temperatures

Figure 11: Diurnal variation of retrieved and expected wavelength correction for a Pandora 121 dataset at Huelva, June 2019. The red line represents the expected wavelength correction due to environmental temperature variation, as characterized in the laboratory. The blue line shows the retrieved correction as output form the spectral fitting. The good match confirms the high quality of the wavelength registration correction in the BSS.



Shift of Serdyuchenko O_3 cross sections

As could be seen from Figure 9 both a wavelength shift in the reference spectrum and the O_3 cross sections can bias O_3temps. Consequently, in order to assess a possible wavelength shift in the O_3 cross-sections, the wavelength registration of the reference spectrum needs to be as precise as possible (as pointed out in the previous paragraph, this required the suggested software improvements.).

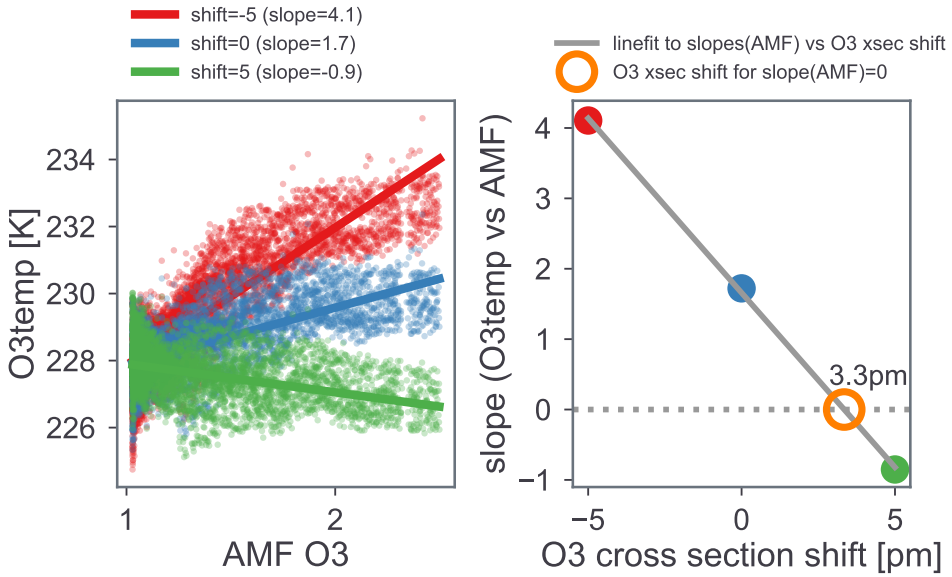
We assume that our tilt-effect "aware"-long window wavelength correction (see section 3.2.4) is adequate for this investigation. Hence, our attempt to reveal a possible wavelength shift of the Serdyuchenko cross section is connected to the valid assumption of O_3temp exhibiting no AMF dependent diurnal variation. We hypothesize that if we would know the actual O_3temp of a particular day, e.g. from O_3 sondes, with all systematic retrieval errors to be sufficiently small, the only reason for a significant variation of a retrieved O_3temp over the day would be a

shift in the O_3 cross sections.

Following this hypothesize we make use of a dataset from Pandora 121, collected during a 10 day Brewer intercomparison campaign in Huelva (Spain), in July 2019. The reference spectrum was picked on a perfect clear sky day at solar noon (to minimize possible AMF dependent errors), where also O_3 sonde data are available. The dataset was processed with three different cross sections: the original ones without wavelength shift, and two more shifted by -5 and +5 pm respectively. As fitting window margins we selected 310 to 330 (our standard window). This also guarantees that we probe only possible shifts from the Fourier Transform spectrometer (below 310 an Echelle grating spectrometer was used).

The left panel in Figure 12 shows the three scenarios displayed as a function of AMF. Red data refer to a -5 pm shift, blue to 0 pm shift and green to a +5 pm shift. The corresponding slopes of a linear fit in O_3temp vs. AMF is given in the figure legend and is again set in relation to the cross section shifts in the right figure panel. This almost perfectly linear correlation clearly shows that the AMF dependency of O_3temp is smallest (= 0, orange circle) when the Serdyuchenko cross sections are shifted by 3.3 pm. This first estimation might be still affected by an insufficient wavelength correction of the reference spectrum.

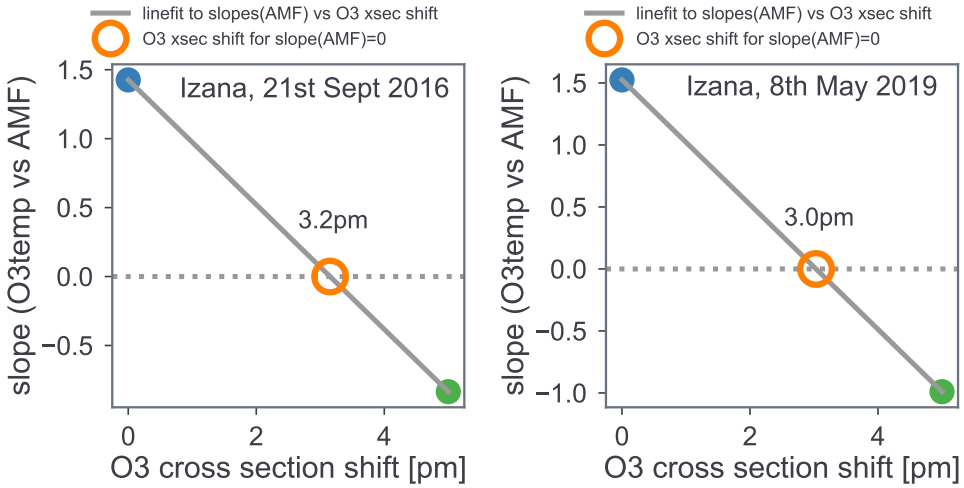
Figure 12: Retrieved O_3temps for 10 days in June 2019 of Pandora 121 at Huelva, Spain. The O_3temp calibration was referenced to a O_3 sonde from the 21st of June. The color-coding on the left panel refers to three scenarios where the utilized Serdyuchenko O_3 cross sections are shifted by -5 (red) and +5 (green) pm and the original one (blue). The O_3temps are given as a function of AMF and a corresponding linear fit is given as solid line. The slopes of the linear fits are shown on the right as a function of wavelength shift. The hypothesized zero dependency on AMF would be reached by a cross section shift of 3.3 pm (orange circle).



Therefore we repeated this test for Pandora 121, however this time using Izana data (ozone sonde data are available too) at two different dates. Figure 13a and 13b show the test results applied for a reference spectrum from the 21st September 2016 and 8th May 2019, respectively. Like for the previous test, a shift of approximately 3 pm was found.

Figure 13: Analysis of Serdyuchenko O_3 cross-sections based on field data from Izana. Same explanation as for Figure 12.

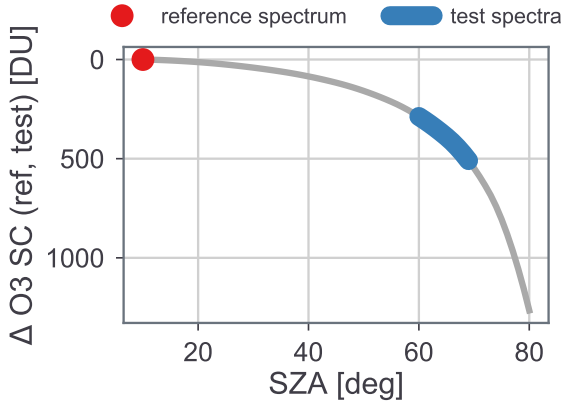
(a) Reference taken at September 21st 2016. **(b)** Reference taken at May 8th 2019.



It should be noted that this test could in principle be applied to other cross section sources too.

Calibration approach for O_3temp

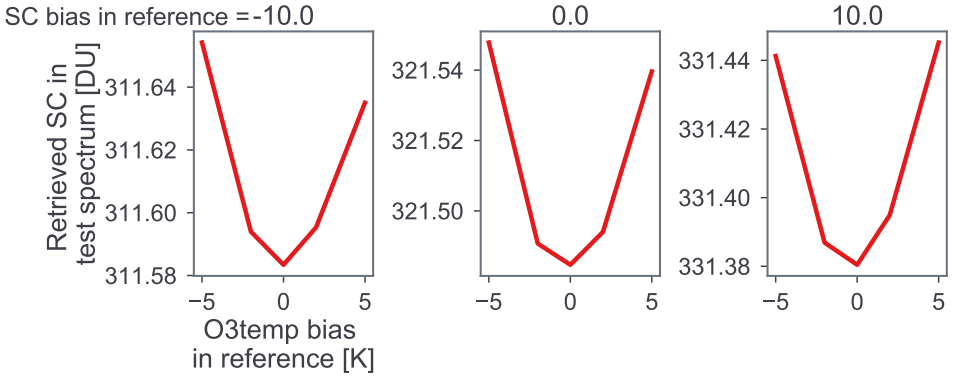
Figure 14: Spectra selection as suggested for the application of the AXC-T. The reference (red dot) is selected at small SZAs. The test spectra are taken from the same day at sufficiently higher SZAs (blue range). The distance is expressed in SC, relative to the reference time.



As introduced earlier, there is in general only little variation of O_3temp over the day for unpolluted conditions, which makes this feature already a very useful help for calibration. However, we want to introduce an independent calibration approach called *Auto Cross Calibration for Temperature* (AXC-T).

AXC-T utilizes a feature emerging from the non-linear temperature dependency of the O_3 cross-sections. Consider a reference spectrum, chosen at a low SZA (red dot in Figure 14). This spectrum is applied to a series of test spectra at sufficiently higher AMFs (blue range in Figure 14, explanation follows later). For this setup, a series of O_3temps are guessed for the reference. If for each of those reference scenarios the retrieved O_3 SC for the test spectra is investigated, one can experience the SC to be biased high if the O_3temp was not guessed correctly. This distinct pattern is shown for simulated data in Figure 15, where the inflection point occurs at the correct O_3temp guess. Further, as shown in the figure, the effect is even independent of the SC assumed in the reference, i.e. the minimum at the correct O_3temp not only arises, when the correct SC is known for the reference (middle panel), but also when it is underestimated by e.g. 10 DU (left panel) or overestimated for the same amount (right panel). Evidently, this method is very robust against a possible error in the reference SC, which is a huge advantage and prompts to perform the O_3temp calibration prior to O_3 calibration itself. Its worth noting that for the calibration of O_3temp, one single clear sky day is sufficient.

Figure 15: Simulated variation of O_3 SCs, retrieved from a test spectrum, for different biases in the O_3temp assumed in the reference spectrum. The figure columns refer to the scenarios where also biases in the O_3 SC in the reference are simulated (left -10 DU bias, middle no bias, right +10 DU bias). The minimum SCs coincide with the correct O_3temp guess in the reference and are robust against incorrect guesses of the SC in the reference.

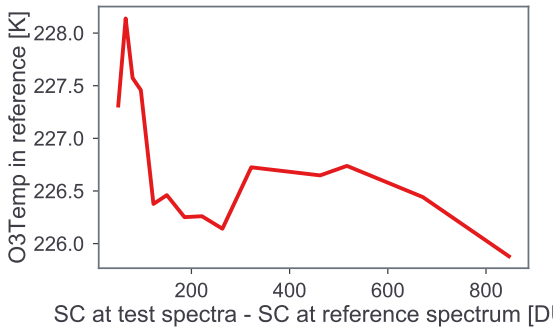


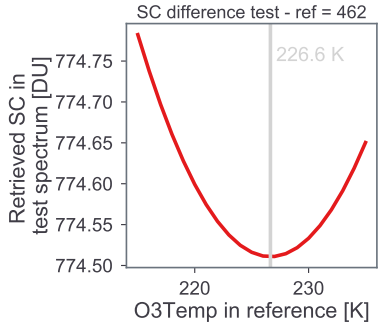
Before we can discuss the needed retrieval settings for AXC-T in more detail, we need to recall the impact of a wrong SpecSL calibration. The lowest laser wavelengths which currently can be measurement in NASA PGN lab, is at 325 nm. Consequently, the SpecSL parameters can be calculated for 325 nm, but need to be extrapolated for wavelengths below. From simulations we can see for a 10 % SpecSL underestimation a positive bias of 2.5 K and for a 10 % overestimation a negative bias of -0.8 K in the O_3temp calibration (inverse effect as for SCs). Uncertainty in the SpecSL extrapolation can hence be a major source of a calibration bias and therefore needs to be considered also in the selection of the retrieval settings.

As reported above, a spectral shift in the O_3 cross-section also biases the O_3temp calibration, but this issue is assumed to be sufficiently solved.

Eventually, two aspects need to be considered for the retrieval setting: the fitting setup and the selected test spectra. For the fitting setup, simulations suggest to rely on low order closure polynomials and, with the aim to mitigate possible stray light issues, to stay away from wavelengths below 305 nm. The test spectra are selected from an AMF range where the AXC-T approach is most stable. This is e.g. shown in Figure 16a for an application in Izana for Pandora 121 for a reference spectrum from the 8th of May 2019. A SC difference of the test spectra to the reference between 300 and 500 DU appears to be appropriate (compare Figure 14) and is in agreement with simulations. The finally picked O_3temp for the reference should be the average over this range. Figure 16b illustrates one AXC-T pattern from this application for a SC difference of about 460 DU.

Figure 16: Example application of the AXC-T for Pandora 121 at Izana from May 8th, 2019.
(a) AXC-T is exemplarily applied for a number of test spectra. The retrieval yields constant O_3temps for test spectra at ΔSC(test, ref) 300 to 500 (compare Figure 14).
(b) One example AXC-T pattern, taken at ΔSC(test, ref) = 462 DU. Compare Figure 16a.





3.5.2 Direct sun total O_3

Already in *Tiefengraber and Cede* [48] we conceptually presented a first implementation of a novel calibration approach which takes advantage of the fact that O_3 can be considered a non-linear absorber. Non-linear not by nature but due to the fact that O_3 absorption and convolution again do not commute owing to sloped and structured absorption features across the instrument bandpass. This is again a consequence of using pre-convoluted cross sections for sake of computational speed. We dub this concept *Auto Cross Calibration for Column* (AXC-C) to reflect the circumstance that it is a self calibration approach (no external data is used). As a reminder, calibration means basically to determine the O_3 SC amount in the reference spectrum at the correct temperature.

Calibration approach for O_3: initial AXC-C concept

This first AXC concept as described in *Tiefengraber and Cede* [48], however, turned out to be not able to deliver unique solutions (no global minimum). Other inves-

tigated approaches, exploiting the fact that different fitting wavelength settings are prone to different non-linearities, turned out to be not stable enough. However they are conceptually working and may possibly applicable to even more non-linear absorbers like water vapor.

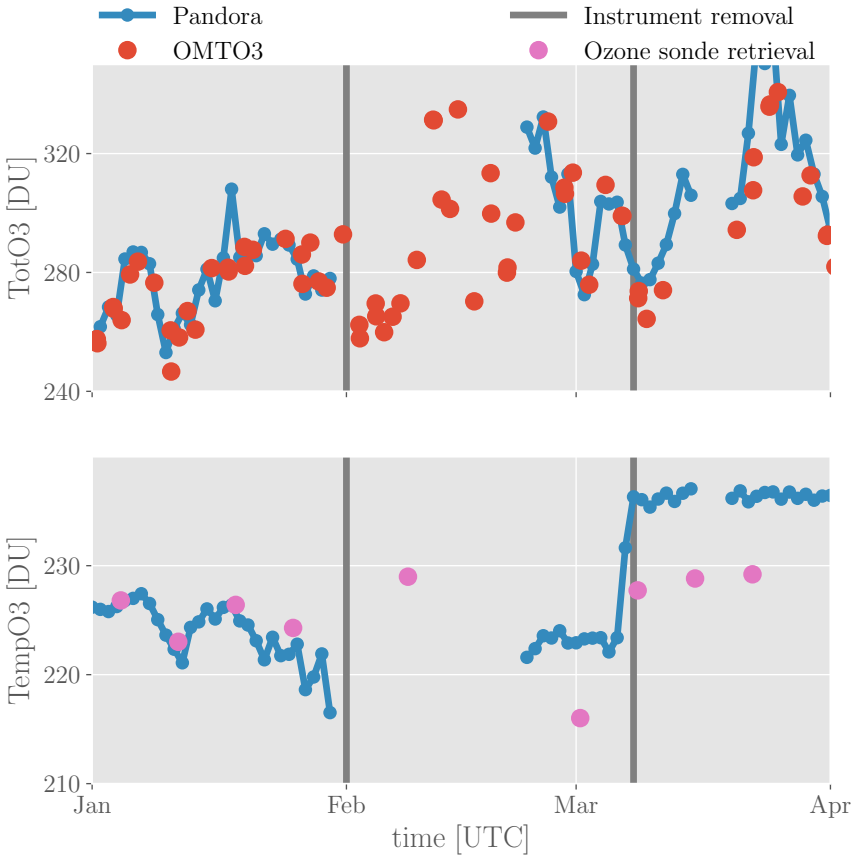
As a consequence, the AXC concept so far still needs "support" by the original Pandora total O_3 product which uses a literature reference. However, first tests of this approach at the station of Izana, Tenerife, for Pandora 121, clearly demonstrated the feasibility and benefit of including O_3temp in the spectral fitting. Figure 17 shows the new retrievals from Pandora as blue line for total O_3 in the top panel, O_3temp in the middle panel and total SO_2 in the bottom panel. Total O_3 from the previous processing is given in pink. For comparison, total O_3 retrievals from OMI (OMTO3) are shown in red and O_3temp retrievals from the weekly ozone sondes launched at the ground station in magenta. The agreement in total O_3 and O_3temp between the datasets is excellent and the very low total SO_2 values for Izana (what could be expected) indicate that biases due to cross correlations are not affecting the data anymore.

O_3temp can be added as a (weak) absorber in the retrieval and is very sensitive on whether the measured reference is valid or not. If, for example, the instrument has to be dismantled for some reason, the reference is probably not valid anymore. This might be undetected when looking at total O_3 alone, but O_3temp immediately reacts and is in general significantly biased. This makes the O_3temp, in addition to having it as an additional atmospheric parameter, an extremely valuable calibration tracer.

An example can be seen in figure 18, where the time series from above (figure 17) is extended by two more months. During that period the instrument had to be dismantled two times (highlighted by gray vertical lines) due to harsh weather conditions. The resulting bias in total O_3 relative to OMI is hardly recognizable, but the “jump” in O_3temp (in particular after the second interaction) is striking. In addition to the bias, O_3temp also shows clearly nonphysical variations over the day, which allows the detection of a calibration change even if the true O_3temp is not known.

Figure 17: Time series of Pandora 121 daily averages of total O_3 (top panel), O_3temp (middle panel) and total SO_2 (lower panel) from May 2016 to Feb 2017 at Izana, Tenerife. The Pandora O_3 product based on the initial AXC is shown in blue. For total O_3 also the current standard product is shown in pink. The OMTO3 O_3 product of OMI is given as red dots and ozone sonde retrievals of O_3temp are displayed in magenta.

Figure 18: Same content as in Figure 17, but without total SO_2. The vertical gray lines highlight the times, when the instrument was dismantled and remounted later (when Pandora the data series continues). The consequential loss of the calibration is not directly recognizable when looking at total O_3, but particularly striking for O_3temp.



Calibration approach for O_3: new AXC-C concept

Due to the situation that the current AXC-C approach still relies on the accuracy of the Pandora total O_3 based on literature reference, we have re-conceptualized AXC-C to circumvent this dependency.

First we need to recall the parameterized implementation of the absorption cross-sections in the BSS, which takes into account a possible non-linearity (as explained before) of certain gases (like e.g. water vapor and O_3) [6]. Per definition, using a non-linear cross-section parameterization demands an absorption free reference spectrum. This is explained in the following: consider the basic Beer-Lambert-Boguer law at one wavelength with only one linear absorber and parameterized according to BSS for two spectra i and k. k shall serve as reference spectrum, A the parameterized cross section (what actually are scaled optical depths in this notation) and qs shall be the relative slant column amount according to the definitions in *Cede* [6].

$$\log(F_i) = \log(F_0) - A \cdot qs_i \tag{1}$$
$$\log(F_k) = \log(F_0) - A \cdot qs_k \tag{2}$$
$$\Rightarrow \log(F_i) = \log(F_k) - A \cdot (qs_i - qs_k) = \log(F_k) - A \cdot \Delta qs \tag{3}$$

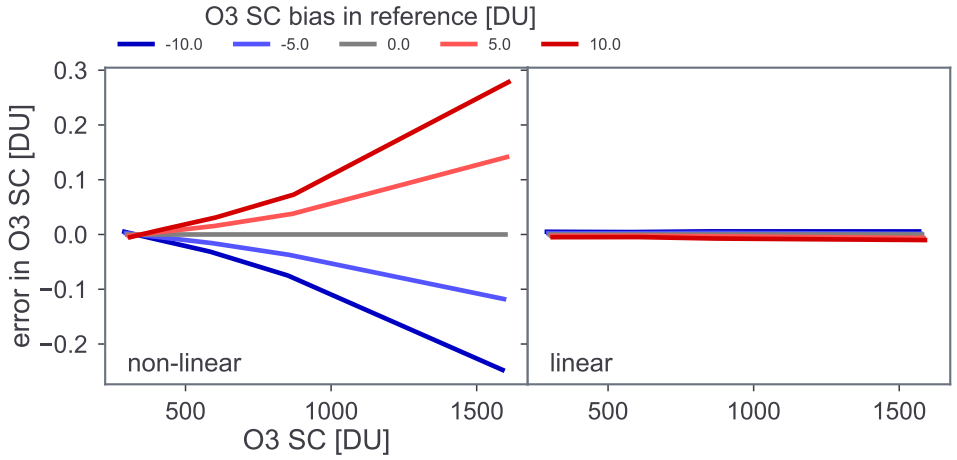
Δqs in (3) is the well known differential slant column density. In case of qs_k being zero (absorption free reference), Δqs would indeed yield absolute values.

Now we consider a non-linear absorber which is again parameterized accordingly. Following the structure as before, spectrum i and k are related via

$$\log(F_i) = \log(F_0) - A \cdot qs_i^{1+B} \tag{4}$$
$$\log(F_k) = \log(F_0) - A \cdot qs_k^{1+B} \tag{5}$$
$$\Rightarrow \log(F_i) = \log(F_k) - A \cdot (qs_i^{1+B} - qs_k^{1+B}) \approx \log(F_k) - A \cdot \Delta qs \tag{6}$$

For this situation Δqs in (6) is only meaningful if qs_k is indeed zero. If this is not the case, a non-linear bias is added. And this circumstance is utilized for the new AXC-C concept. Figure 19 illustrates this bias for different guesses of O_3 SC in the reference as a function of O_3 SC. No impact of a wrong guess is discernible for a linear algorithm (right panel), but a non-linear algorithm "reacts" (left panel).

Figure 19: Simulated error in the retrieved O_3 SC when using a synthetic reference spectrum (at SZA = 0 deg). Different guesses on the SC amount in the reference are color-coded. For a non-linear algorithm (left figure panel), the O_3 SC dependency (or AMF or SZA) is only constant when the SC was guessed correctly. Linear algorithms (right figure panel) are insensitive to this effect (explanation in the text).



This effect is AMF dependent (as evident from figure 19), because the non-linearity is more pronounced for larger AMFs. However, the effect is generally rather small for weakly non-linear absorber like O_3 and makes it necessary to remove sufficiently well all other AMF dependent disturbances from the data.

A conclusion from above is that in principle a retrieval algorithm set up in the linear way is insensitive to errors in the reference SC value, but a non-linear algorithm is sensitive. Therefore, if one would retrieve O_3 columns both ways (linear and non-linear), the results would only agree if the reference is indeed absorption free.

On top of that one needs also to consider that O_3 is indeed a non-linear absorber and consequently applying a linear algorithm leads to an AMF dependent error. This error would overwhelm the wanted feature from an incorrect SC guess. However, this error can be "modeled" by comparing the O_3 retrievals for the non-linear and linear algorithm when using a literature reference, because the literature reference is per definition absorption free.

Taking all this into account one can define a double difference (or delta) for the retrieved absolute slant column qs^* like

$$\Delta^{\mathrm{L}} qs^*_{\mathrm{syn}} = qs^*_{\mathrm{nl,syn}} - qs^*_{\mathrm{l,syn}} \tag{7}$$

$$\Delta^{\mathrm{L}} qs^*_{\mathrm{lit}} = qs^*_{\mathrm{nl,lit}} - qs^*_{\mathrm{l,lit}} \tag{8}$$

$$\Delta^{\mathrm{R}}\Delta^{\mathrm{L}} qs^* = \Delta^{\mathrm{L}} qs^*_{\mathrm{syn}} - \Delta^{\mathrm{L}} qs^*_{\mathrm{lit}} \stackrel{!}{=} 0 \qquad \text{for all AMFs.} \tag{9}$$

With

$qs^*_{\mathrm{nl,syn}}$	absolute slant column retrieved from non-linear algorithm and synthetic reference.
$qs^*_{\mathrm{l,syn}}$	absolute slant column retrieved from linear algorithm and synthetic reference.
$qs^*_{\mathrm{nl,ext}}$	absolute slant column retrieved from non-linear algorithm and literature reference.
$qs^*_{\mathrm{l,ext}}$	absolute slant column retrieved from linear algorithm and literature reference.
$\Delta^{\mathrm{L}} qs^*_{\mathrm{syn}}$	Linearity (L) delta based on synthetic reference.
$\Delta^{\mathrm{L}} qs^*_{\mathrm{ext}}$	Linearity delta based on literature reference.
$\Delta^{\mathrm{R}}\Delta^{\mathrm{L}} qs^*$	Delta of applied references (R) and linearity algorithms.

Expression

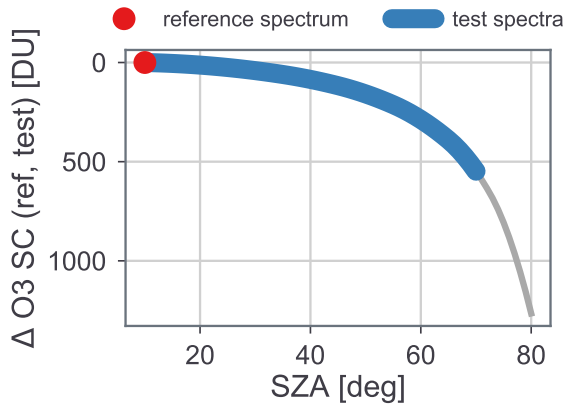
(7) contains the "calibration sensitivity" factor $qs^*_{\mathrm{nl,syn}}$

(8) is needed to correct for the intrinsic difference between qs^*_{nl} and qs^*_{l}

(9) can only be fulfilled if the synthetic reference spectrum is indeed absorption free.

$\Delta^{\mathrm{R}}\Delta^{\mathrm{L}} qs^*$ is calculated for a number of test spectra at AMFs below ≈ 3 (blue line in Figure 20), whereas the references is the same spectrum as for O_3temp calibration (red dot in Figure 20). Again, one single clear sky day is sufficient for the O_3 calibration.

Figure 20: Spectra selection as suggested for the application of the AXC-C. The reference (red dot) is selected at small SZAs. The test spectra are taken from the same day up to about AMF 3 (blue range). The distance is expressed in SC, relative to the reference time.



This double delta approach comes with the advantage that possible biases in qs^*, which are constant in AMF, cancel out. Nevertheless, in order to meet the above mentioned "basic requirements", which is to sufficiently remove all other AMF dependent factors, key is the best possible characterization of the instrument - in particular SpecSL characteristics.

For the AXC-C, SpecSL comes into play again. Like for O3temp calibration (AXC-T), incorrect SpecSL characterization biases the estimated O_3 value in the reference. From simulations we can see for a 10 % SpecSL underestimation a negative bias of -2.5 DU and for a 10 % overestimation a positive bias of 0.2 DU.

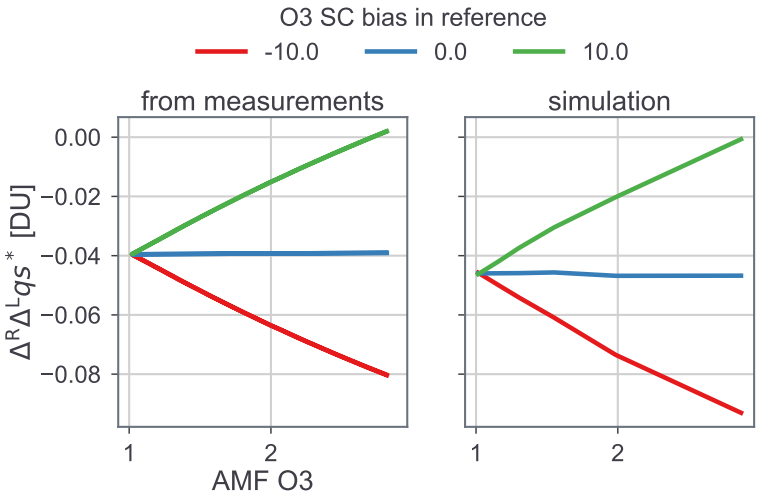
Next to the impact of SpecSL, a potential bias in the estimation of O_3temp needs to be assessed. First, O_3temp is used to scale the O_3 cross sections for the retrieval of qs^* and hence it has to be assumed that O_3temp is constant over day. This is unavoidable, because O_3temp can not be retrieved using a literature reference. Second, O_3temp is also used to create the synthetic reference. From simulations we see a 2 K error in the O_3temp estimation to result in about 1 DU error for the O_3 SC calibration.

The spectral fitting setup for the AXC-C relies on lowest order closure polynomials. This is key because particularly retrievals in the UV spectral range can suffer from background polynomials of higher order, because they are able to explain partly the baseline slope of the O_3 absorption in the fitting.

An application of AXC-C is displayed in the Figure 21 for Pandora 121 at Izana (left figure panel). For this application, the reference was taken from the 8th of May 2019 around noon and have been applied to test spectra data from the same day. The different lines in the figure represent different SC guesses relative to the "known" value (red -10 DU bias, blue correct guess and green 10 DU bias). As explained above, the correct guess yields to vanishing AMF dependencies of $\Delta^{\mathrm{R}}\Delta^{\mathrm{L}}qs^*$. This is accompanied by the simulated $\Delta^{\mathrm{R}}\Delta^{\mathrm{L}}qs^*$, shown in the right figure panel. The pattern is almost identical. Note that the small (and unimportant) offset in the intercept stems from the difficulty to properly simulate the intrinsic differences between the literature reference spectrum and Pandora spectra.

Further it's worth mentioning that in general three SC guesses are sufficient. The slope of a linear regression in $\Delta^{\mathrm{R}}\Delta^{\mathrm{L}}qs^*$ as a function of AMF is almost perfectly linearly connected for different (rather close) SC guesses and can be therefore safely inter- or extrapolated (not shown here).

Figure 21: Example application of the AXC-C applied to measurement for Pandora 121 at Izana, May 8th, 2019, (left figure panel). There is a very high agreement to AXC-C based on simulations (right figure panel).



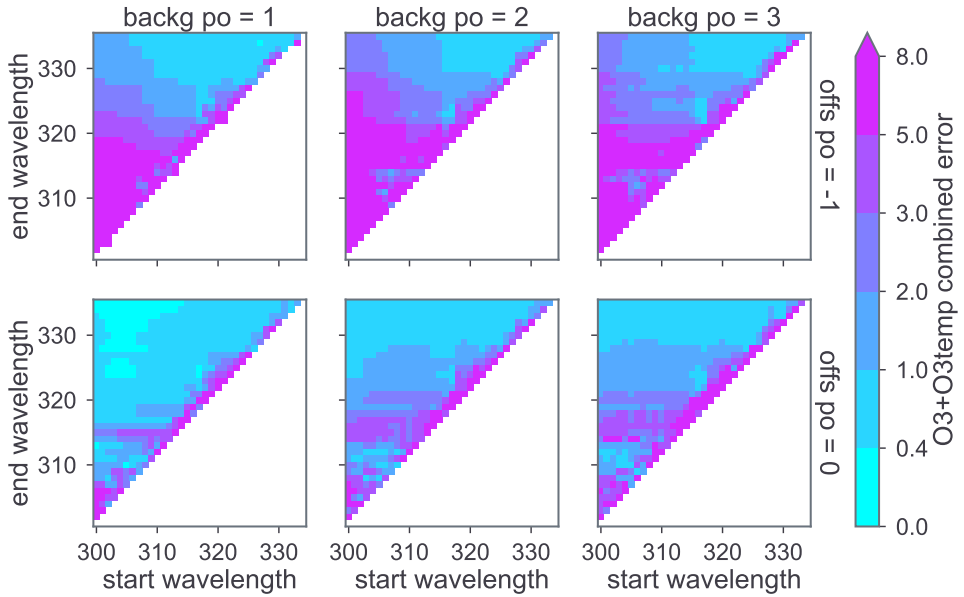
We want to point out that the fitting setup for the calibration and the final data product is different. The setup for the calibration is geared towards minimizing systematic AMF differences between the usage of different reference types. The

latter one has the goal to be least prone to systematic biases and, at the same time, as representative for as much atmospheric conditions as possible (AMFs, aerosol conditions, etc.).

3.5.3 Fitting setup for O_3 and O_3temp

The optimal fitting setup should minimize systematic errors for O_3 columns and O_3temp, since both parameters are retrieved simultaneously. Hence the combined error was simulated for different fitting setups and is shown in Figure 23 as wavelength matrix plot. Tested are different sets of start (x-axis) and end (y-axis) wavelengths of the fitting window as well as background polynomials (columns) and offset polynomials (rows). The simulation was done for several SZAs. Displayed are the average values between 0 and 80 deg.

Figure 23: Simulated combined error for the retrieval of O_3 and O_3temp as a function of fitting window start (x-axis) and end (y-axis) wavelength, as well as the order of the background polynomial (columns) and offset polynomial (rows).



A feature what is frequently seen for retrievals in the UV and slightly contradicts common practice of the DOAS community, is that higher order background polynomials (figure columns) tend to increase the error. Although this certainly does not hold true in general, for direct sun retrievals in the UV, higher order background polynomials are able to "explain" SpecSL features at the cost of O_3 correctness.

Further, the analysis suggests fitting starting wavelengths between 301 and 310 nm and ending wavelengths between 325 to 335 nm. Due to the known issue regarding SpecSL, a conservative selection is favorable. As a consequence of the temperature fitting, an offset polynomial of order 0 appears to be more stable as well.

3.5.4 Example applications

In the previous section the underlying calibration procedures (AXC-T+C) for the new O_3 product was outlined. Now we want to investigate some example applications.

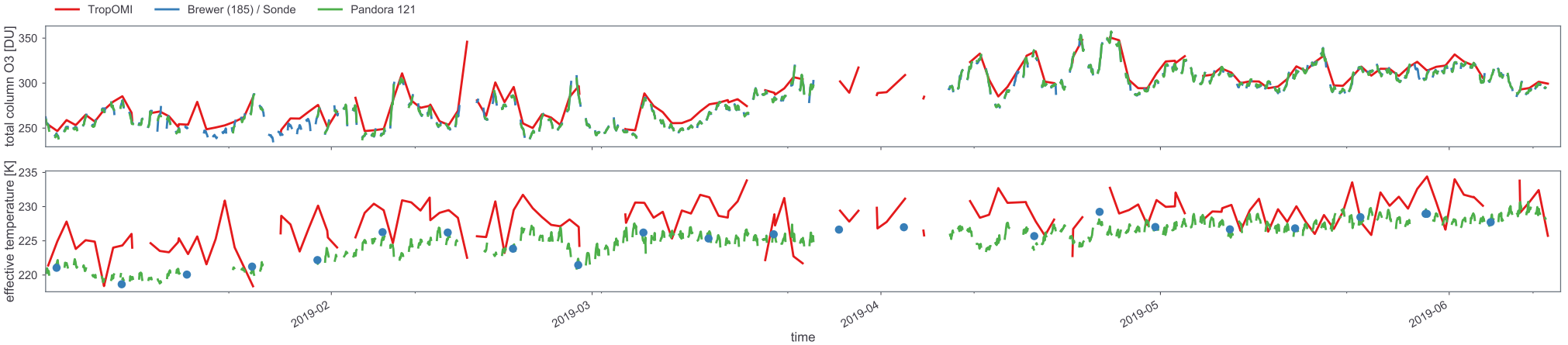
Izana, Tenerife, half-year 2019

Izana represents a perfect location for O_3 data comparison for two reasons. Firstly, Izana hosts the European reference Brewer triad and secondly, ozone sondes are launched once a week. We make use of these datasets, which have been thankfully processed and provided by Alberto Redondas.

Further, the Pandora at Izana experienced a rigorous SpecSL characterization with a tunable laser at BTP in Braunschweig. This reduces the main uncertainties for AXC-C+T.

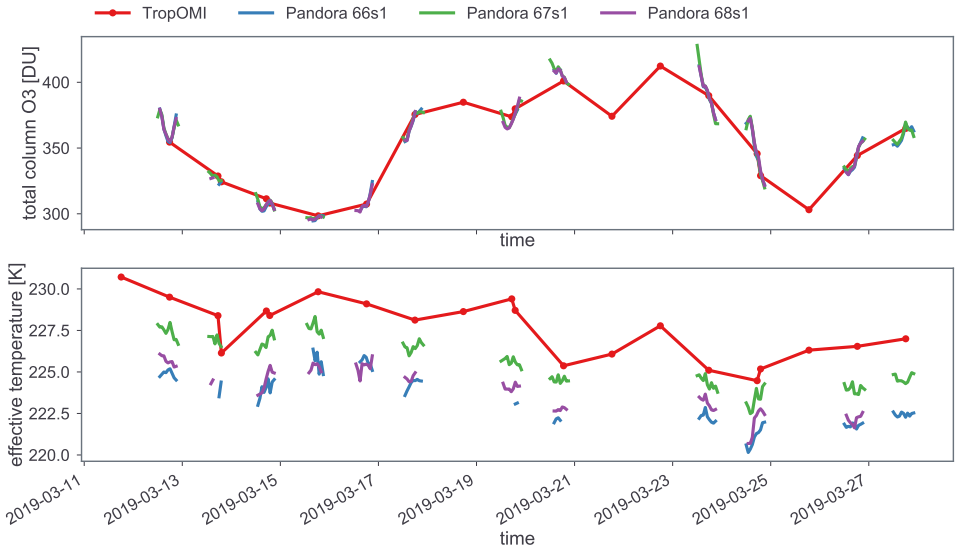
Figure 22 shows data from January to June 2019 for Pandora 121 in green, total columns in the top figure panel and the effective temperatures in the bottom panel. For reference, O_3 total columns from the Brewer #185 and O_3temp retrievals from ozone sonde are shown in blue. The comparison is complemented by the TropOMI (reprocessed) offline O_3 product, which gives beside the column values also estimations about the effective temperature [45]. We see a very good agreement to the reference datasets both for O_3 columns and O_3temp. The O_3temp from satellite is retrieved as slant information and also uses the TOMS O_3 profile climatology as a-priori information. The difference in observation geometry between the satellite and a direct sun ground based instrument makes it not straight forward to interpret the observed differences in O_3temp.

Figure 22: Diurnal variation of total column O_3 (top panel) and effective O_3 temperature (bottom panel) at Izana (Tenerife, Spain), 2019. Retrievals from Pandora 121 are shown in green. For reference, total columns from the reference Brewer #185 and temperature retrievals from sondes are shown in blue. Complementary, TropOMI (reprocessed) offline O_3 data products are displayed as red lines.



Goddard Space Flight Center, Maryland, March 2019

Figure 24: Diurnal variation of total column O_3 (top panel) and effective O_3 temperature (bottom panel) at Goddard Space Flight Center (MD, USA), March 2019. Retrievals from Pandora 66, 67 and 68 are shown in blue, green and purple, respectively. Complementary, TropOMI (reprocessed) offline O_3 data products are displayed as red lines.



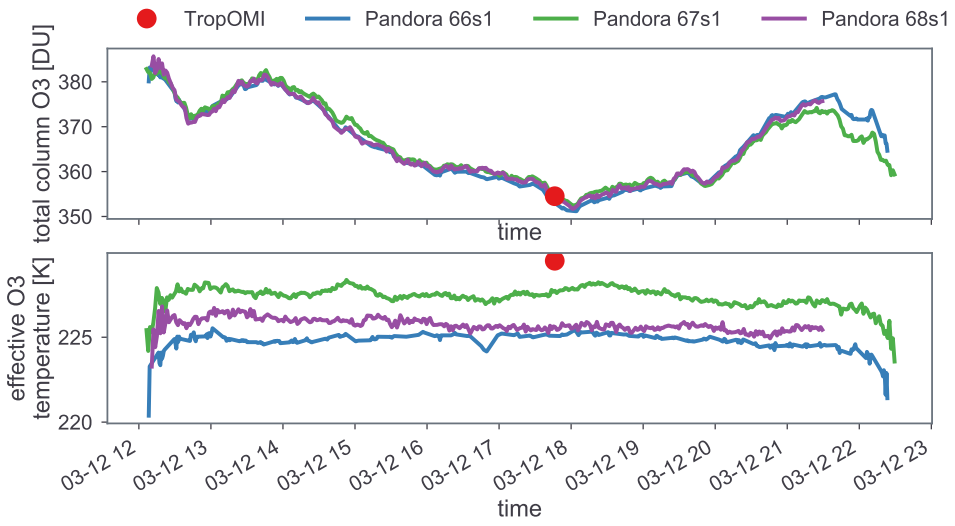
Goddard, being one of PGNs main laboratory calibration sites, offers the advantage of having quite often several instruments gathered for field testing. Therefore we picked a period in March 2019, where Pandora 66, 67 and 68 were active. Also, because for those instruments the majority of available lasers, in particular the 325 nm laser, have been measured in the laboratory.

Since all three Pandoras have been independently calibrated, this setup also allows us to give a rudimentary estimate about the quality of the AXC method. As in Figure 22, the top panel of Figure 24 displays O_3 columns and the bottom panel O_3temp. Shown is Pandora 66 in blue, 67 in green and 68 in purple. Again the TropOMI (reprocessed) offline O_3 data products are overlain in red.

From a first glance one can evidence a very close agreement for the O_3 columns among the Pandoras. For O_3temps, Pandora 66 and 68 agree adequately, but Pandora 67 is off by about 2 K. A close up for the 12th of March 2019 is shown in Figure 25.

Considering uncertainties in SpecSL characterization as a main driver for O_3temp biases, the observed discrepancies most likely are related to that. The impact on O_3 calibration is small, due to AXC-C being sufficiently robust against smaller O_3temp biases.

Figure 25: Same explanation as for Figure 24. Shown is a close up of March 12th, 2019.



In terms of satellite validation, the larger spread for O_3temp does not necessarily pose a problem. Direct validation of O_3temp from satellite is certainly interesting but probably its interpretation not totally straight forward due the difference in the sampled air masses. As pointed out in the motivation of O_3temp retrieval, first and foremost O_3temp serves as tracer for the validity of a calibration. However, once we have a better understanding about the O_3temp variation over the day, valuable implications for tropospheric O_3 assessments are likely.

3.5.5 Operational O_3 retrieval products

As part of the developments of the BSSv1.8 (see Table 1), the TOMS version 8 temperature and ozone profile climatology [16] was incorporated. First, this allows us a latitudinal and time dependent estimate of the effective ozone height needed for the AMF calculation. Second, the O_3temp itself can be estimated as a function of time, latitude and O_3 column amount as well.

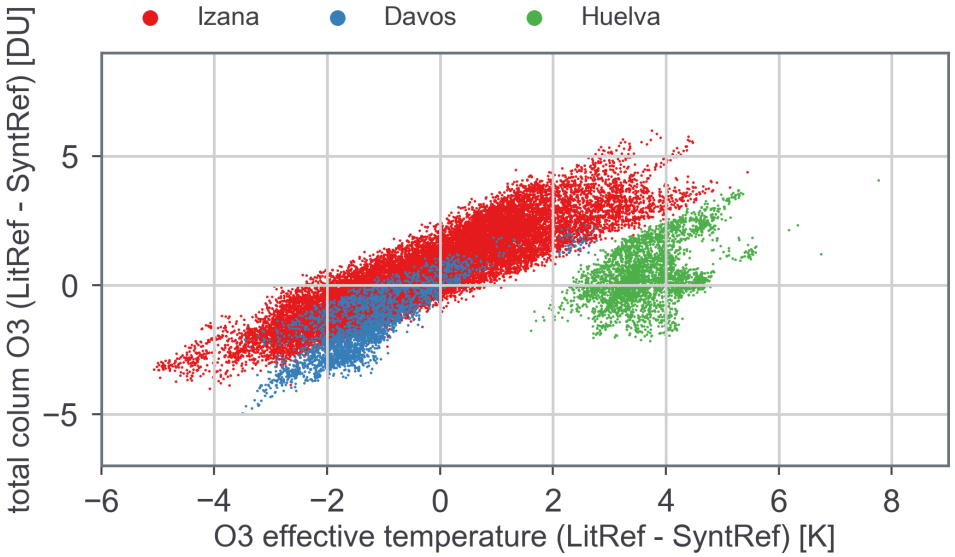
This improvement, together with the mentioned effective absolute calibration for instruments (see Table 1) and a fitting setup optimization, makes the utilization of a retrieval product based on the literature reference feasible again (LitRef).

The retrieval of O_3temp, similar to the retrieval of other "minor" absorbers, is rather sensitive to instrumental issues like bad pointing (combined with a bad FOV) or spectral features. Considering the large number of Pandoras also at different operational quality and calibration states, makes the AXC based O_3/O_3temp product (SyntRef) not applicable in all cases. In theses cases the O_3 product based on the literature reference is a very good approximation already. In particular for low latitude stations, with rather small O_3temp dynamics.

In fact, first applications turned out that the difference between the O_3 product based on AXC and the (new) O_3 product using the literature reference can often be explained by the difference in the O_3temps (retrieved vs. climatology) only. In Figure 26 the difference in O_3 columns of both products as a function of the difference between the climatological and retrieved O_3temp is shown for three examples (Izana, Davos, Huelva). The column difference is seen to be within $\pm$5 DU and clearly depends on the differences between O_3temps.

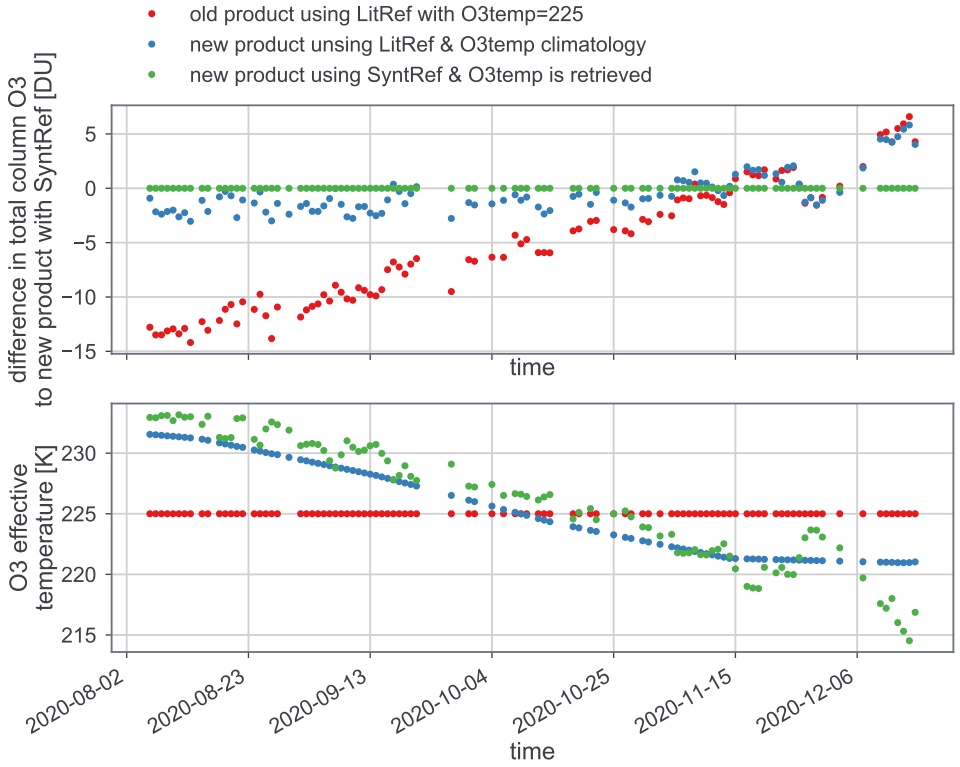
Obviously, the relation does not hold true for all datasets like for Huelva. It is however not clear yet why. A potential bias in the O_3temp retrieval from Pandora 121s1 in Huelva could explain this offset, but O_3temp compares very well with Ozonesondes at Huelva. This needs further investigations.

Figure 26: Difference of Pandora O_3 column retrievals between using a literature reference and O_3temp climatology (LitRef) and a synthetic reference with O_3temp fitting (SyntRef). This difference is shown as a function of the difference in both O_3temp estimations (climatology vs. retrieved) for datasets from Izana (red), Davos (blue) and Huelva (green). Except for Huelva, the there is a clear relation that the difference in O_3 columns of both products is driven by the difference in the O_3temp estimation.



Example time series (Davos, 2020) of the two new O_3 products and the former product are shown in Figure 27 as diurnal mean values. The seasonal and small scale variations in O_3temp, as retrieved by the AXC based O_3 product are obvious (bottom panel, green dots). The impact on the O_3 retrieval is striking, with bias amplitudes reaching 15 DU for the old O_3 product and, as mentioned, up to 5 DU for the new O_3 product based on the literature reference.

Figure 27: Comparison between the old O_3 product (red) and the two new O_3 products using either a synthetic reference and O_3temp fitting (SyntRef, green) or a literature reference and a O_3temp climatology (LitRef, blue). The top panel shows the absolute O_3 column difference to SyntRef. The bottom panel shows the retrieved (SyntRef), climatological (LitRef) or fixed (old product) O_3temps.



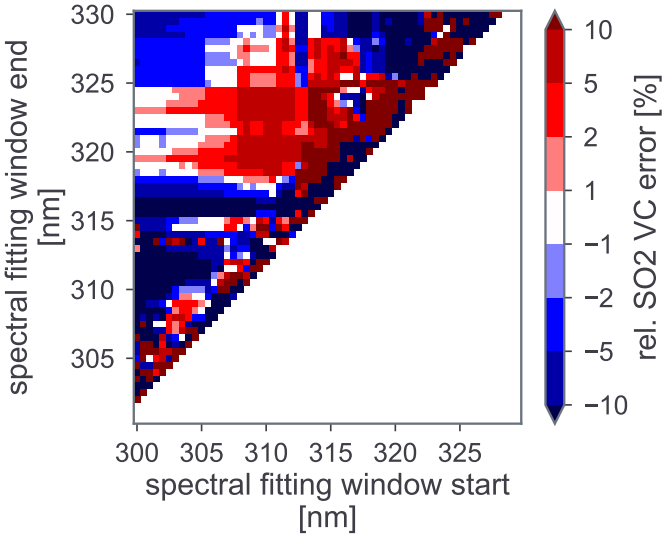
3.6 Direct sun total SO_2

The strongest absorber in the spectral region of SO_2 absorption is O_3 and also features related to O_3temp. This is why it is of importance to have O_3 absorption properly corrected in a synthetic reference spectrum which is used for total SO_2 retrievals.

3.6.1 Fitting setup

As suggested by sensitivity studies based on SIMPLE, the optimal fitting setup for SO_2 uses a fitting window between 305 and 325 nm. The corresponding wavelength matrix plot is shown in figure 28, where relative SO_2 slant column (SC) errors in percent are displayed for different sets of start (x-axis) and end (y-axis) wavelengths of the fitting window. The simulation was done for several SZAs, but shown is the average between 20 and 70 deg. One can recognize that a stable spectral region (white area) is around the selected fitting window limits. Comparable analysis have been performed likewise to determine most suitable closure polynomial orders (not shown here). The exact fitting setup can be seen in the PGN data products readme document [8] and is downloadable from the PGN web page.

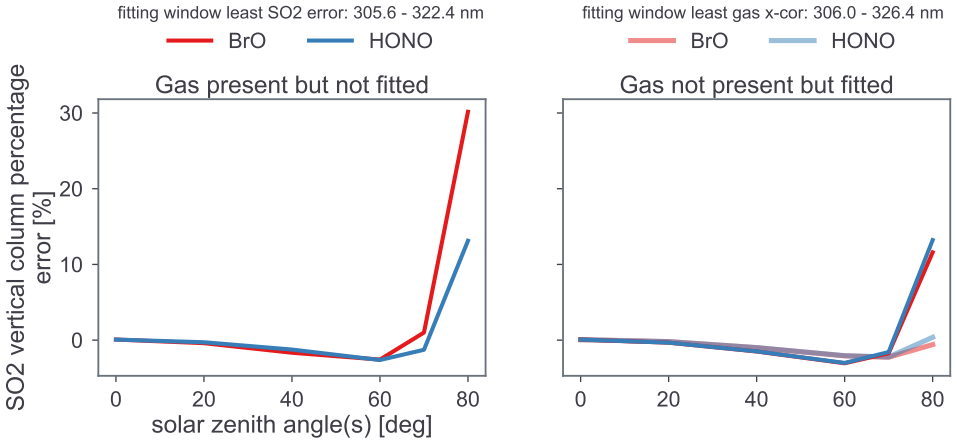
Figure 28: Simulated percentage error of SO_2 SC as a function of fitting window start (x-axis) and end (y-axis) wavelengths. Data are averaged between SZA 20 and 70 deg. A stable fitting window region is around 305 to 325 nm.



In particular for a weak absorber (like SO_2) potential cross-correlations to other gases what might be included in the fitting setup needs to be checked. It needs to be answered whether it is better to include all potential gases even if they might not even be present in the atmosphere all the time or all stations (like e.g. BrO, with highest concentration in the polar regions). For the SO_2 fitting window, the potential additional gases would be BrO and HONO. The simulated impact of adding the gases or not is shown in Figure 29 as a function of the SZA with the gases color coded. The left figure panel deals with the situation when a gas is present in the atmosphere but not fitted. The right figure panel instead, faces the situation when the gas is fitted but not present. The simulation was carried out with the fitting setup yielding the smallest SO_2 errors, but this window appears to have a rather strong HONO and in particular BrO dependency at higher SZAs. Since the impact is even worse if BrO is not fitted but present, it is suggested to include BrO and HONO in the fitting window in any case. This effect considered, we could develop a more appropriate fitting window with reduced impact of BrO and HONO (faint red and blue in the right figure panel).

Please note that the simulation assumes constant gas amounts (also in the reference) and that BrO and HONO have not been calibrated in this simulation (since they are no PGN products so far). This is why the error at lower SZAs vanishes. If the gases would have been calibrated the error at small SZAs would still be almost zero.

Figure 29: Simulated SO_2 retrieval error when either BrO or HONO (colors) is present in the atmosphere but not fitted (left panel), or not present but fitted (right panel).

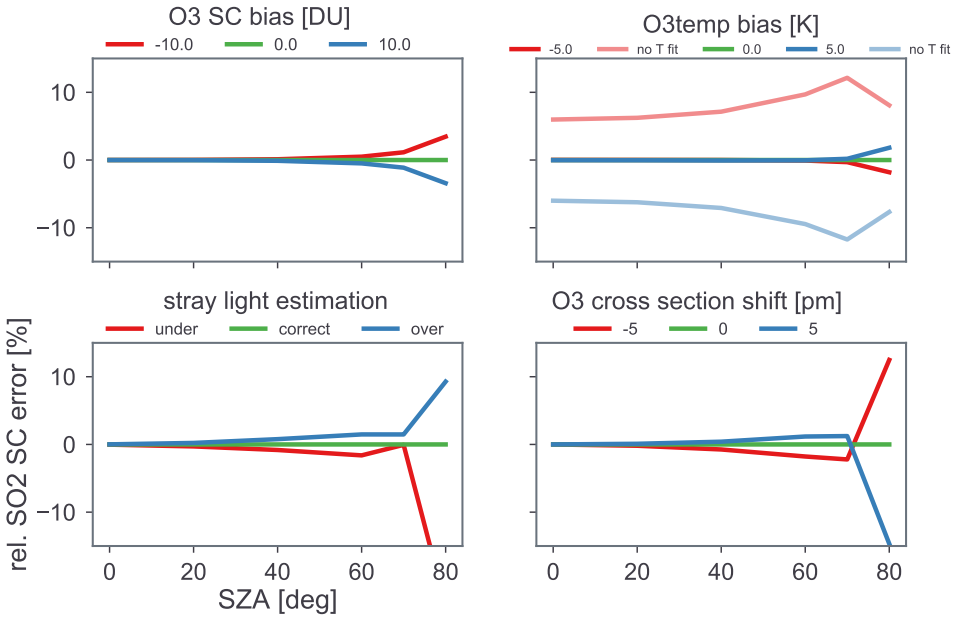


3.6.2 O_3 cross-sensitivity

In the previous sections we have identified topics which could possibly limit total SO_2 accuracy (O_3 (+temperature) absorption, SpecSL, O_3 cross section displacement). Their percentage impact was quantified using SIMPLE simulations and is summarized in figure 31. Four parameters are tested within a rather narrow range, reflecting the estimated accuracy limits of the current state and are referenced to the unbiased case. Those parameters are: a bias in O_3 SC in the reference (top left), the corresponding O_3temp (top right), SpecSL (bottom left) and a possible O_3 cross section shift (bottom right). Please note that for each tested parameter, all other ones were not biased.

Generally speaking, all tested parameters show a sufficiently small impact on total SO_2 retrieval accuracy for SZA < 70 deg. We only want to point out the importance of allowing the fitting of O_3temp for total SO_2 retrieval (figure panel on top right): when the O_3temp was estimated incorrectly by e.g. 5 K, the impact on total SO_2 is still very small as long as O_3temp fitting is allowed (dark blue line). If for the same situation temperature fitting is not enabled, the emerging spectral features are wrongly explained by SO_2 absorption which leads to biases (light blue line). For this case even the actual atmospheric temperature is correct in the simulation, only the O_3temp in the reference is biased. If also the atmospheric temperature differs (which would probably the case in reality), this effect even amplifies. To our knowledge, current SO_2 algorithms for ground based instruments do not take into account variable O_3temp.

Figure 31: Simulated percentage error of SO_2 SCs as a function of a bias in the O_3 SC in the reference (top left), O_3temp in the reference (top right), SpecSL correction (bottom left) and a shift in the O_3 cross sections (bottom right). For the O_3temp case, also the effect is shown when no O_3temp fitting is included in the fitting (faint colors).



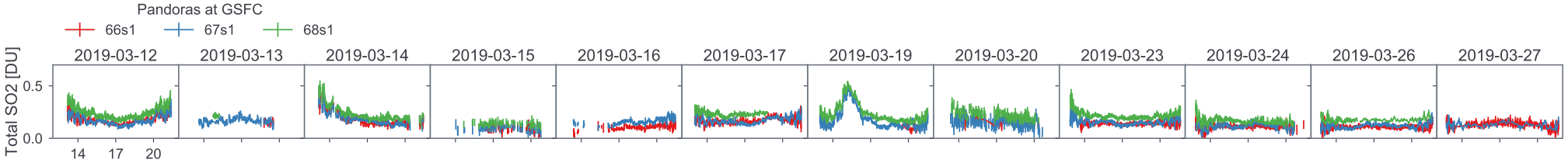
3.6.3 Calibration of SO_2

The vertical distribution and chemistry of SO_2 advertise again the use of the MLE for determining the SO_2 SC amount in the reference. In contrast to NO_2, where a quasi constant (stratospheric) background is present, we do not expect to have a significant background for SO_2.

3.6.4 Example application

Proper validation of total SO_2 is not possible at the current state due to the absence of external SO_2 reference datasets. Instead we make use of three co-located Pandoras at GSFC in March 2019 to compare total SO_2 from those instruments (Pandora 66, 67 and 68). All three datasets have been calibrated independently and are shown in Figure 30 with errorbars. While Pandora 66 and 67 agree extremely

Figure 30: Diurnal variation of total SO_2 at GSFC for Pandora 66 (red), 67 (blue) and 68 (green). Datasets are calibrated independently, as explained in the text.



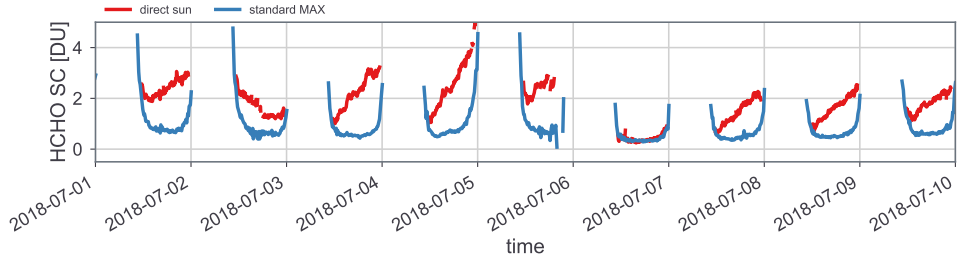
well, number 68 is slightly higher. This is most likely connected to uncertainties in the wavelength grid of the reference as the formerly mentioned tilt-effect is not well considered yet. In general we see the majority of the time rather low and constant total SO_2 values over the day - as one would expect. On the 19th in the morning hours, a SO_2 event is well captured by all instruments. Overall we can see a remarkably low point-to-point variation with slightly enhanced uncertainties for higher AMFs.

3.7 Direct sun total HCHO

Although direct sun total HCHO is an operational PGN data product, it is not available for all instruments for the hole time series. This is due to the fact that up to mid 2019 all produced Pandora sensor head units included pieces made out of Delrin. Delrin significantly outgases HCHO. This circumstance was not clear until intensive investigations, including comparison to multi axis (MAX)-DOAS data, revealed this situation[6]. Figure 32 is a representative picture showing one example HCHO SC amount of Pandora 108 in Egbert. Instead of showing the expected AMF driven U-shaped diurnal variation of SC amounts, the direct sun HCHO contains AMF independent HCHO in the sensor head what is building up due to heating over the day. During cooler phases this outgasing is reduced and parts of HCHO are reabsorbed (e.g. on the 6th of July this was the case due to strong winds). This lead to significant seasonal biases between total HCHO from direct sun and MAX DOAS based retrievals, which almost no difference in winter and huge differences during summer. A dedicated publication was submitted already to AMT and is currently in review [44].

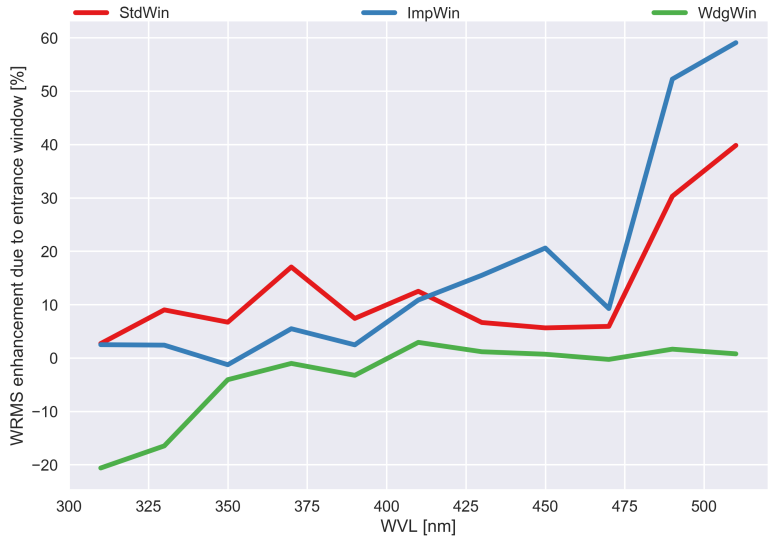
[6]MAX-DOAS based retrievals usually use for each measurement sequence one reference spectrum, e.g. the zenith spectrum. By this the excess HCHO amount in the head is always canceled out. In turn, Pandora direct sun total algorithms usually use on reference spectrum for a long time period.

Figure 32: Diurnal variations of HCHO SC, retrieved from direct sun Pandora data (red) and Pandora sky data (blue). HCHO based on sky data (from the preliminary test product) show the expected U-shape variation, as the excess HCHO in the sensor head is canceled out by selecting a new reference for each measurement sequence. In contrast, direct sun retrievals of HCHO use only one reference, hence the excess HCHO in not canceling out. Note that on the 6th of July, cooling due to strong winds helped to suppress HCHO outgasing.



Since summer 2019 all new Pandoras use Nylon instead of Delrin gears and existing units are modified accordingly.

Figure 33: Comparison of the RMS enhancement by using an entrance window compared to no window. The results of two different coated windows are show in red and blue, those for a wedged window in green. Obviously using a wedged window introduces no additional spectral signals.



3.7.1 Unwanted spectral signal from entrance window

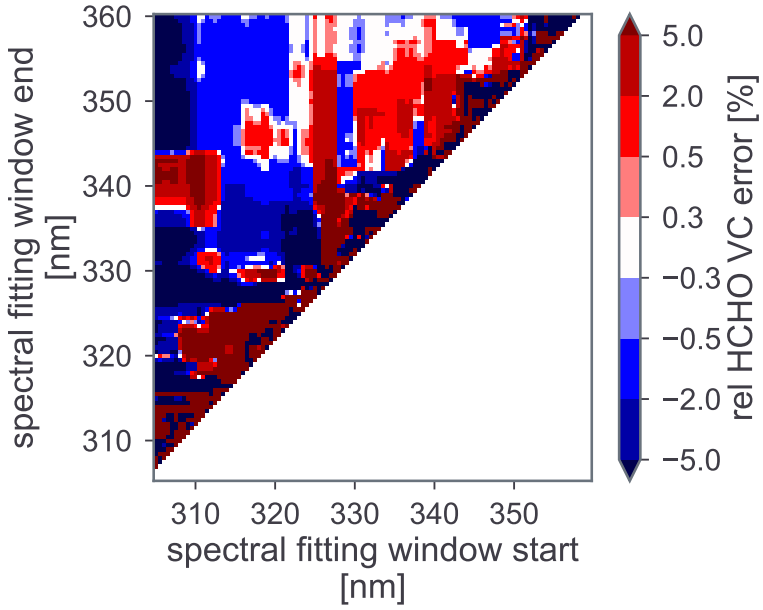
Although current Pandora sensor heads are equipped with coated entrance windows to reduce interference effects ("unwanted" spectral signal USS see e.g. *Tiefengraber and Cede* [47]), these coatings seem to be not efficient to fully eliminate these effects. Hence have been introducing wedged windows instead of the coated ones, because due to the wedge side, the cavity reflected beam is simply baffled away. Studies carried out at Izana could show that wedged windows indeed perform comparable to having no entrance window at all. This is shown in figure 33, where the enhancement of the spectral fitting residual RMS is shown compared to having no window. Two different coated windows (red and blue curve) still show an enhancement of the RMS, while the wedged window performs like no window (the negative values are probably an artifact of the analysis). Based on the successful tests at Izana, almost all Pandora entrance windows have been replaced in the mean while and this issues is considered solved.

3.7.2 Fitting setup

Sensitivity studies based on SIMPLE have been carried out in order to estimate the optimal fitting setup for direct sun total HCHO retrievals. The simulation incorporates all characterized instrumental features and their (accurate) correction. The corresponding wavelength matrix plot is shown in Figure 34, where relative HCHO VC errors in percent are displayed for different sets of start (x-axis) and end (y-axis) wavelengths of the fitting window. The simulation was done for several SZAs. Displayed are the average values between 0 and 80 deg.

The analysis suggests several areas of theoretically low HCHO retrieval errors, usually reaching to the end of the HCHO cross-sections (about 360 nm). We usually emphasize larger fitting windows. This is why we select the smallest feasible start wavelength around 322 nm. The exact fitting setup can be seen in the PGN data products readme document [8] and is downloadable from the PGN web page.

Figure 34: Simulated relative HCHO VC retrieval error as a function of fitting window start (x-axis) and end (y-axis) wavelength. Data correspond to averages between SZA 0 and 80 deg.

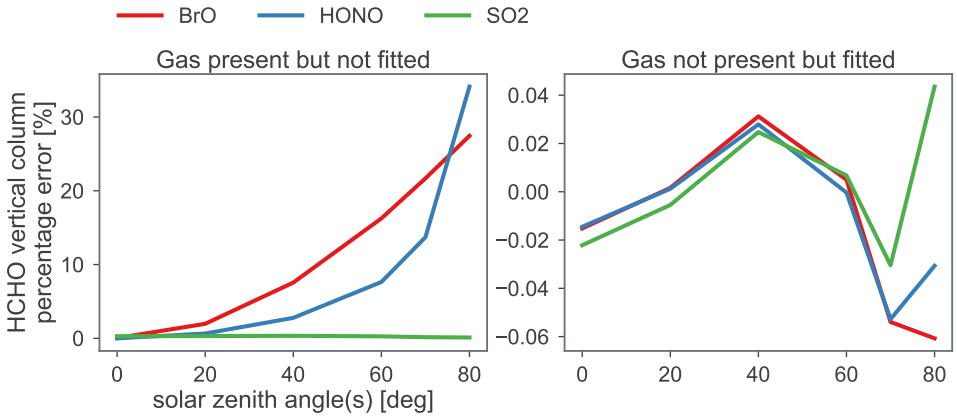


Similar as done for the SO_2 retrieval setup, potential cross-correlation to other gases

what might be included in the fitting setup needs to be checked. It needs to be answered whether it is better to include all potential gases even if they might not even be present in the atmosphere all the time or all stations (like e.g. BrO, with highest concentration in the polar regions). For the HCHO fitting window, the potential additional gases would be BrO, HONO and SO_2. The simulated impact of adding the gases or not is shown in Figure 35 as a function of the SZA with the gases color coded. The left figure panel deals with the situation when a gas is present in the atmosphere but not fitted. The right figure panel instead, faces the situation when the gas is fitted but not present. The results clearly show that in any case it is better to include all possible absorbers in the spectral fitting.

Please note that the simulation assumes constant gas amounts (also in the reference) and that BrO and HONO have not been calibrated in this simulation (since they are no products so far). This is why the error at lower SZAs vanishes. If the gases would have been calibrated also the error at small SZAs would increase drastically to 30 % (10 %) for BrO (HONO).

Figure 35: Simulated HCHO retrieval error when either BrO, HONO or SO_2 (colors) is present in the atmosphere but not fitted (left panel), or not present but fitted (right panel).



3.7.3 HCHO calibration

Up to now, the MLE is the standard calibration technique used for HCHO calibration. However, the following particularity makes the application of MLE challenging for HCHO calibration: Due to the connection of HCHO chemistry to biogenic reactions (see e.g. *Spinei et al.* [43]), a typical diurnal HCHO pattern commonly reaches a maximum shortly after solar noon. A wrong guess of the reference slant column in the MLE also leads to a comparable AMF dependent diurnal bias. Thus HCHO variability can interfere with the MLE.

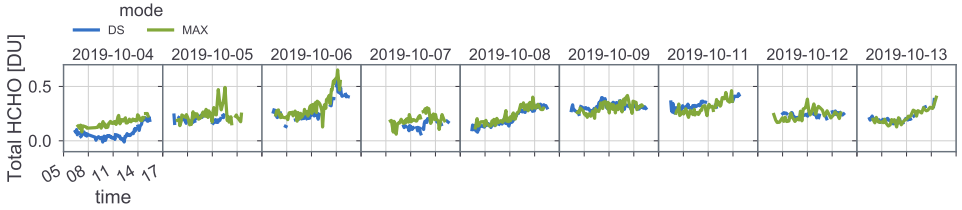
This aspect brings to mind, that again the variable, tropospheric part of the column is the main problem. This suggests to use the introduced E-MLE (see section 3.3) as the calibration approach, which was implemented in BSS v1.8.

3.7.4 Example application

Delrin free sensor heads have been available only lately and hence continuous time series are sparse. Further, due to the lack of reference data sets, we consider it reasonable to compare total HCHO from direct sun with tropospheric HCHO, retrieved from the Pandora MAX mode. The latter though, is mainly available for European and South-American Pandoras, since the Pandora from the US (with longer Delrin-free time series in principle) haven't operated the MAX mode of Pandora so far. Consequently we end up with one possible dataset to look at: Pandora 138 in Rome.

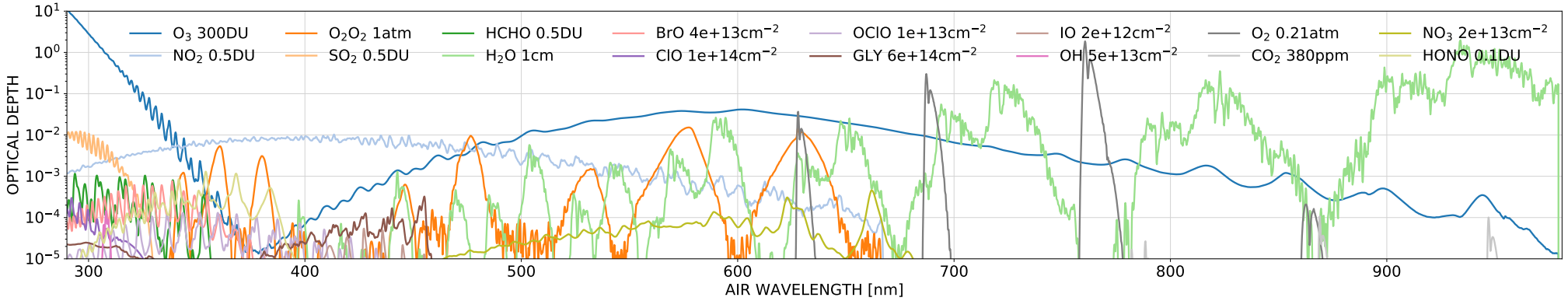
For Pandora 138, the general agreement between direct sun (DS) total (blue lines in Figure 37) and MAX based tropospheric (green lines in Figure 37) HCHO retrievals is good. However, every now and then the direct sun data of Pandora 138 seem to be affected by an issues which creates an AMF dependent bias, as seen on the first and fourth day of the time series in the figure.

Figure 37: Diurnal variation of HCHO retrievals for Pandora 138 in Rome, operated in direct sun mode (DS = blue) and multi axis mode (MAX = green).



We know form Pandora 138 that there have been tracker issues from time to time (still the old model is in use). Bad pointing might be able to explain this. This issue needs further investigation and most importantly other Pandora datasets have to be studied as well.

Figure 36: Optical depths of the assumed vertical standard amount of trace gases included in the spectral fitting library of the Blick software suite.



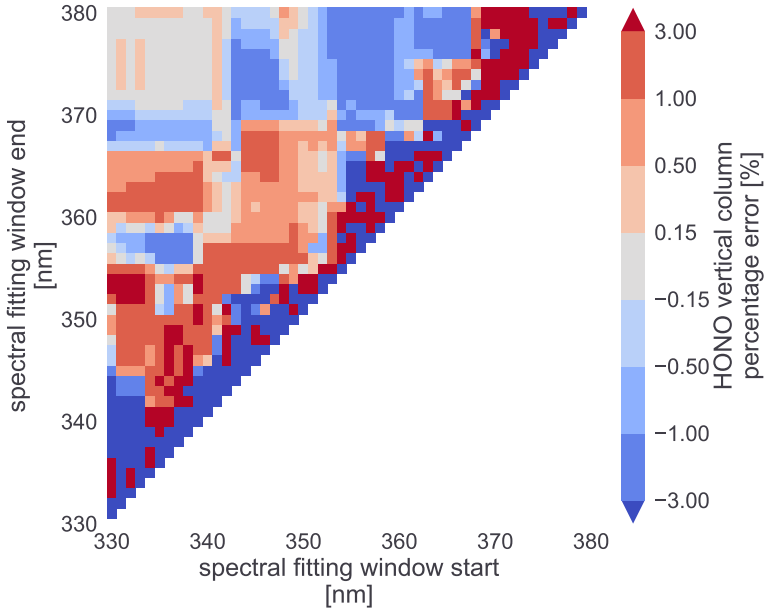
3.8 Direct sun total HONO

Nitrous acid (HONO) is a weekly absorbing gas with ODs staying well below 10^{-3} for typical atmospheric concentrations. Its spectroscopic retrieval hence demands very well characterized and stable DOAS instruments. Since this is in particular difficult to achieve for DS data (where usually a Fraunhofer reference spectrum is used for long time periods), current (and historic) literature only reports HONO retrievals based on long path and MAX DOAS observations (where the Fraunhofer reference is adjusted frequently).

As we will experience later, for good reason HONO is not yet an official PGN data product.

3.8.1 Fitting setup

Figure 38: Simulated HONO VC retrieval error at a SZA of 70 deg as a function of fitting window start (x-axis) and end (y-axis) wavelength. The error is given relative to the true values.



As commonly done for PGN data products, sensitivity simulations based on SIMPLE have been conducted to estimate the best performing fitting setup for direct sun total HONO retrievals. The simulation includes all instrumental characteristics (as spectrally dependent slit function parameters) and their (accurate) correction. The corresponding wavelength matrix plot is shown in Figure 38, where HONO VC errors (relative to the truth) in percent are displayed for different sets of start (x-axis) and end (y-axis) wavelengths of the fitting window. Shown are the results for a SZA of 70 deg, while the reference spectrum was taken at SZA of 1 deg. Areas in gray indicate fitting windows with the smallest systematic errors.

The preferable fitting window starts between 334 to 341 nm and ends between 371 to 379 nm. Although HONO absorption features extend up to 400 nm, we have limited the simulations to 380 nm - which commonly coincides with the throughput limit of the U340 band pass filter used for this type of retrievals (= in the UV).

For a HONO inter-comparison study using CINDI-2 data [52], the fitting window 335 to 373 nm was used. We have decided to favour 340 to 377 nm, which features a vanishing SZA dependent error by still very small overall bias. As for other minor absorber, a smoothing polynomial of order 4 (5 parameters) is selected to potentially also compensate smooth instrumental spectral features (e.g. from misalignment). And, following the conclusions from previous studies (shown above), all potential absorber in the fitting window (except OClO) are considered.

3.8.2 Calibration

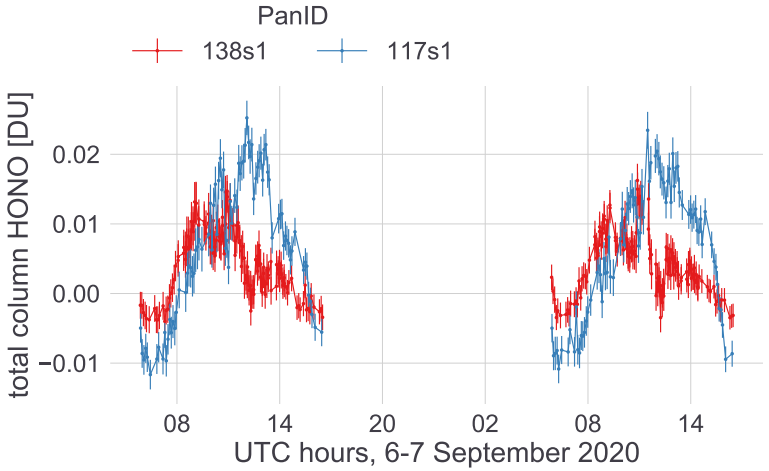
HONO is (optically speaking) a very linear absorber and further is expected to be predominantly present in the boundary layer. Consequently, the MLE is an appropriate method for HONO calibration (that is to quantify the HONO load in the reference spectrum).

3.8.3 Example applications

As shown previously on synthetic data, HONO retrieval settings with negligible residual errors for DS observations are feasible. Unfortunately, several tests on field data could not confirm these findings yet, as Figure 39 exemplarily shows for two colocated Pandoras (117s1, 138s1) in Rome, Sapienza university. This example datasets do not agree in their diurnal variation and also show negative values at higher AMFs (hence they are not correlated with calibration). Although not all tested datasets go below zero, the enhanced disagreement between co-located datasets persist.

The disagreement between different instruments clearly states that the retrieval is distorted depending to different extent by different instruments. As summarized in section 5, we suspect an etalon structure induced by the sandwiched filter setup used for this type of retrieval being mostly responsible for the disagreement. This issue needs to be solved before this data product can be finalized.

Figure 39: Direct sun total columns of HONO as retrieved from Pandora 117s1 and 138s1 during September (6,7) 2020 in Rome Sapienza university. The disagreement between both datasets cannot be explained yet and is most likely related to unaccounted instruments features.



3.9 Direct sun total CHOCHO

Glyoxal (CHOCHO) is a weekly absorbing gas with typical ODs even below HCHO and HONO (hence « 10^{-3}). As it is retrieved in the spectral region around 450 nm, limitations outlined in section 5 impact the retrieval of CHOCHO ot an extent what prevents us from finalizing the data product at that stage.

CHOCHO would be of particular interest if considered together with HCHO. As CHOCHO and HCHO follow different pathways in the oxidation chain of VOCs, their ratio enables more insight on the speciation of precursor VOC [e.g. 31].

3.10 Direct sun total O_2O_2

Direct sun total O_2O_2 is currently not an operational data product, but being tested as part of routine quality control to detect instrumental changes which require a new calibration. O_2O_2 exhibits the favorable character of being only dependent on surface pressure and temperature and is hence expected to show a rather smooth and constant variation over the day. This characteristic makes O_2O_2 columns perfectly suited as QC parameter. Further the low variability of O_2O_2 makes the MLE a suitable approach for calibration.

For Pandora S1 instruments, it is reasonable to use retrieval fitting windows covering the two well-defined bands in the UV (around 360 nm) and VIS (around 480 nm) range (Figure 36). Either fitting setup further includes the strongest absorbers present in the window. The retrieval parameters are also summarized in Table 3.

Since Rayleigh scattering is subtracted before the spectral fitting, the background polynomial (BPol) only needs to capture aerosol extinction. Hence for a perfect instrument, a linear BPol would be sufficient to close the retrieval for this short fitting windows. Consequently if a higher order BPol is needed to yield the expected smooth diurnal variation, this polynomial can be interpreted as the responsible instrumental spectral feature and hence indicates instrumental changes (typical values of needed BPols are given in Table 3 in brackets).

Table 3: Fitting setups for direct sun total column O_2O_2 windows. Column 'BPol' reports the background polynomial order. For a perfect instrument a first order BPol would be sufficient. However, due to instrumental changes, higher order BPols are needed (in brackets).

	Wavelength [nm]	Fitted gases	BPol
UV	346-370	O_2O_2, O_3, NO_2, BRO, HCHO, HONO	1 (2)
VIS	455-490	O_2O_2, O_3, NO_2, H20, I2, OIO	1 (3)

3.10.1 Example application

Direct sun O_2O_2 is being tested for the fitting setups as previously described for a testing instrument located at Wakkerstroom, South Africa.

Figure 40 shows the reference day for BPol 1,2,3. Clearly, applying BPol=1 shows a very strong variation along the day, which is significantly reduced by higher orders which are able to capture disturbing spectral features (not shown). For this UV fitting window, BPol=2 seems already sufficient since there is not much difference compared to BPol=3. This is also visible in the wrms, where BPol=2,3 show a similar, and expected U-shape over the day (Figure 41).

An example timeseries comparing the retrieved O_2O_2 columns from the two bands is shown in Figure 42. The 1st order polynomials illustrate large discrepancies in the daily shapes for both bands, while with BPol=3, the daily shapes show a good agreement.

If an instrument is well-calibrated and not affected by (spectral) instrumental features, low order polynomials should be sufficient and, in particular, O_2O_2 columns from two different bands should yield similar total columns. BPols leading to similar O_2O_2 columns represent, as already mentioned, the smooth spectral distortion in the system. Stringing the polynomials from the different O_2O_2 column retrievals together, would therefore allow to reconstruct the smooth spectral distortion over a larger spectral range (considering also the other possible O_2O_2 bands).

Figure 40: Direct sun O_2O_2 retrieved for the UV fitting window on the reference day 2020-07-14.

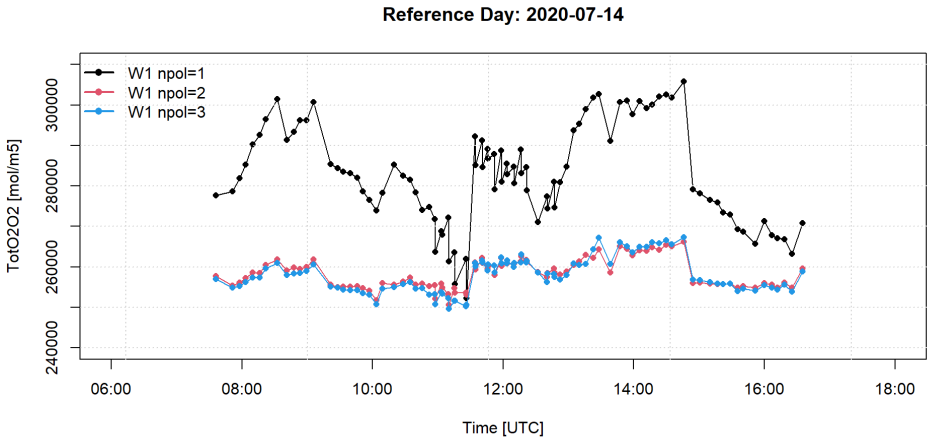


Figure 41: Direct sun O_2O_2 wrms retrieved for the UV fitting window on the reference day 2020-07-14.

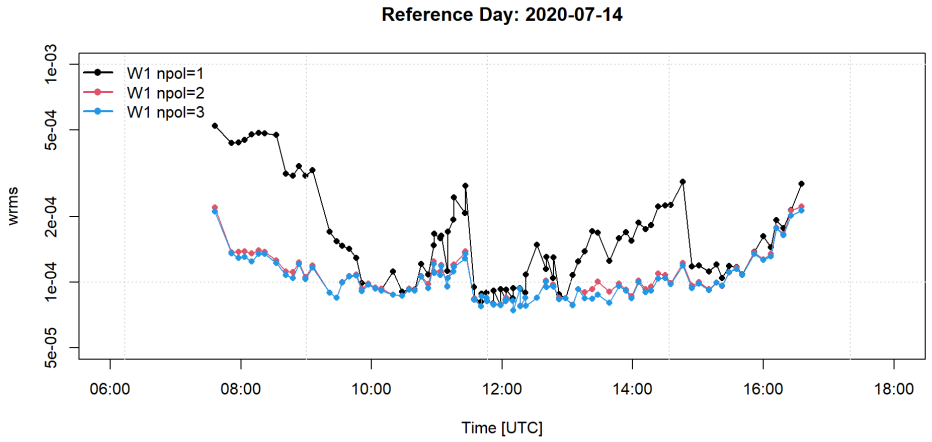


Figure 42: Direct sun O_2O_2 timeseries retrieved for the UV and VIS fitting window for Wakkerstroom, South Africa from 2020-05-02 to 2020-05-05.

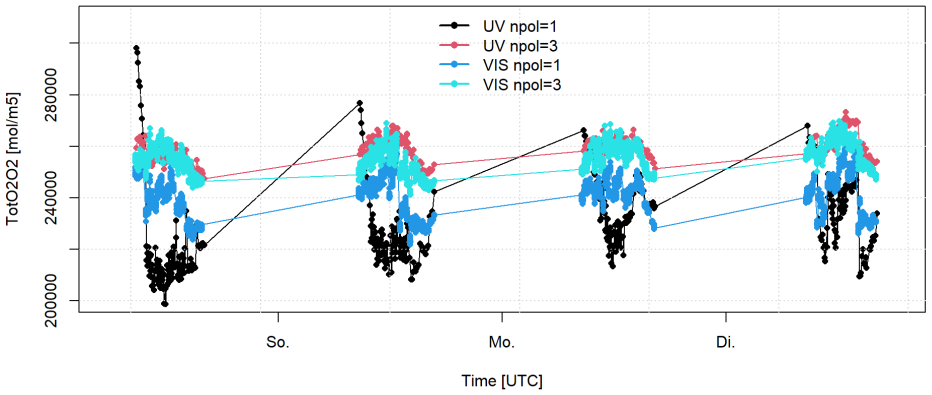
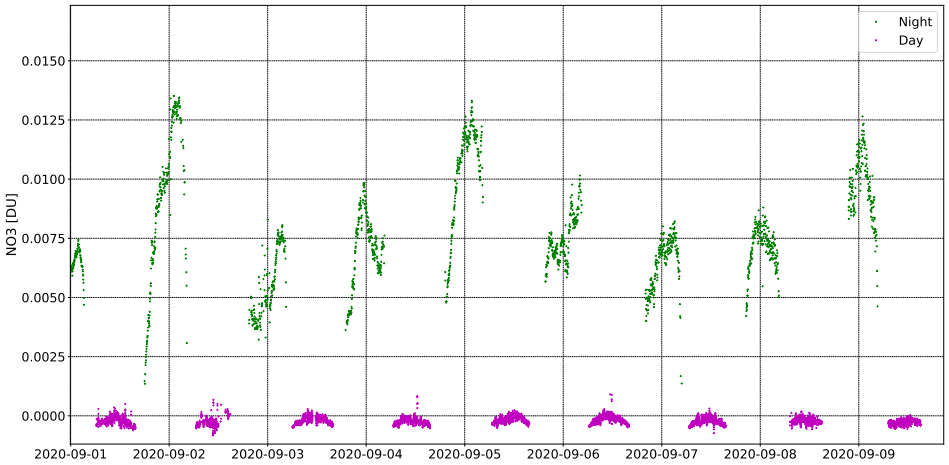


Figure 43: Direct sun/moon total NO_3 column for P138s2 at Rome Sapienza



3.11 Direct sun/moon total NO_3

NO_3 is an important nocturnal absorbers as it affects the nitrogen cycle and secondary formation of pollutants. NO_3 is formed primarily by the reaction of O_3 with NO_2 (*Atkinson et al.* [14], *Geyer et al.* [26], *Matsumoto and Tanaka* [35], *Russell et al.* [40]).

NO_3 can be detected using the Pandora 2S system (600 - 700 nm range), from it's 2 peaks at 620 and 660 nm. A first retrieval setup has been applied in the spectral range of 600 to 685 nm, covering the two peaks. An MLE, calibration method is performed with synthetic reference spectrum from solar measurements. Additional gases that have to be considered are for fitting in this range are O_3, H_2O, O_2O_2, O_2 and NO_2. For this retrieval, the molecular scattering is subtracted and a smoothing polynomial of 4 is used.

Figure 43 illustrates direct sun total NO_3 columns at day and night times from spectrometer 2 data of Pandora 138, located at Rome Sapienza. The top panel is the NO_3 columns retrieved using MLE calibrated for moon using the sun reference and bottom panel is the NO_3 calibrated with sun using sun reference.

Figure 43 illustrates direct sun total NO_3 columns from spectrometer 2 data of Pandora 138, located at Rome Sapienza. In next trials, the NO_3 retrievals using different fitting windows and moon reference will be performed.

Figure 44: Direct sun/moon total NO_3 column for P138s2 at Rome Sapienza

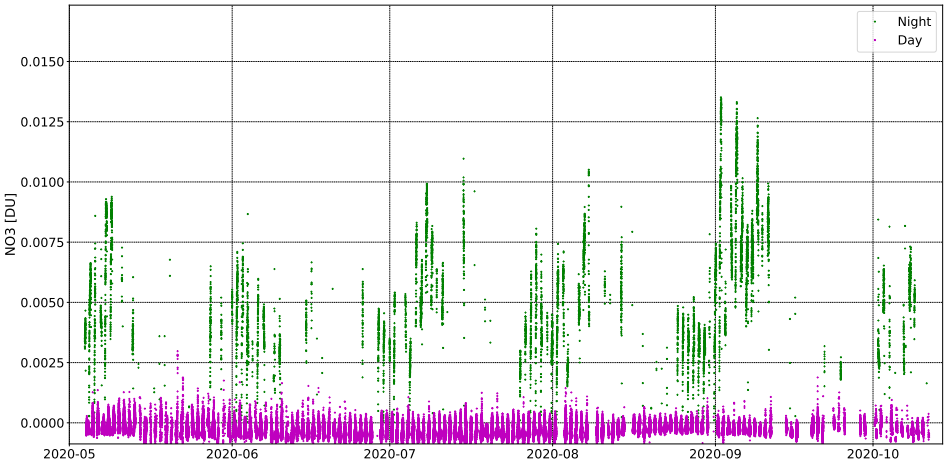


Figure 44 illustrates direct sun total NO_3 columns for a longer period from spectrometer 2 data of Pandora 138, located at Rome Sapienza.

3.12 Direct sun total H_2O

3.12.1 Feasibility concept

H_2O has multiple spectral ranges where it could be retrieved, and it should be possible to retrieve the same H_2O amount for all of them. The OD for each range is increasing with increasing wavelength, with a simultaneous decrease of minor absorbers that have to be taken into account. Therefore, ranges in the upper visible range of a Pandora 2S would be preferred with ODs for standard concentrations well above 1e-2. However, these ranges also reveal a strong temperature dependence in their OD as wavelength increases, which is balancing the positive impact of a stronger signal. Additionally, H_2O is characterized by highly structured cross-sections which require non-linear OD methods.

Conclusively, the feasibility concept for direct sun H_2O retrievals is to retrieve total column amount H_2O from different spectral ranges with optimized fitting windows.

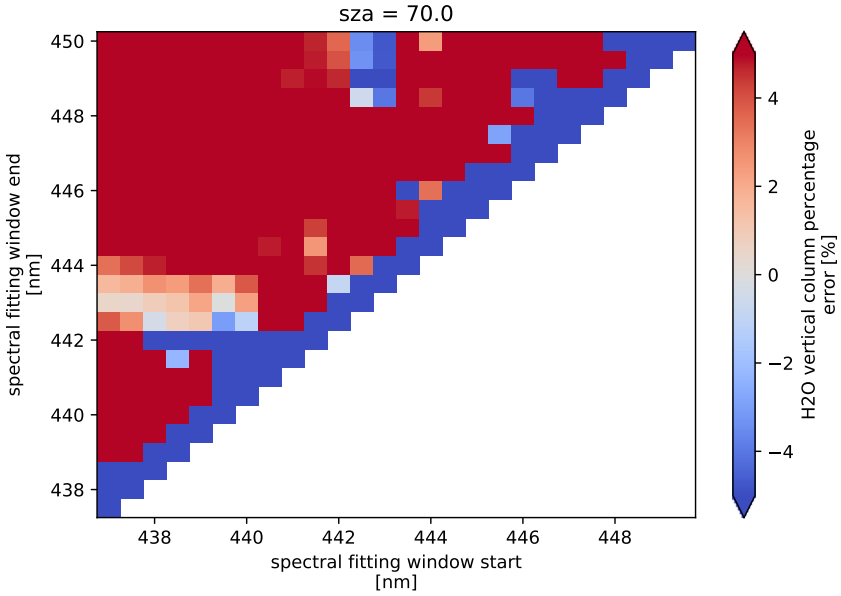
3.12.2 Fitting setup

In the following, 5 different spectral ranges are tested. The SIMPLE simulations are illustrated for SZA at 70 deg with a reference spectra at 1 deg, where the matrix plot shows the fitting start-wavelength on the x-axis and the fitting end-wavelength on the y-axis. Gray areas highlight regions that with the smallest potential percentage error in the retrieved total vertical column amount.

In all simulations and tested fitting windows the polynomials for offset,wavelength change, and resolution change were set to 0.

437-450 nm (A, Figure 45) The gases considered in this range are O_3, NO_2, H_2O, O_2O_2. The simulation suggests a fitting window from 437-443 nm, which is rather small and capturing only the highest H_2O peak. Moreover, the simulation suggests a second order smoothing polynomial.

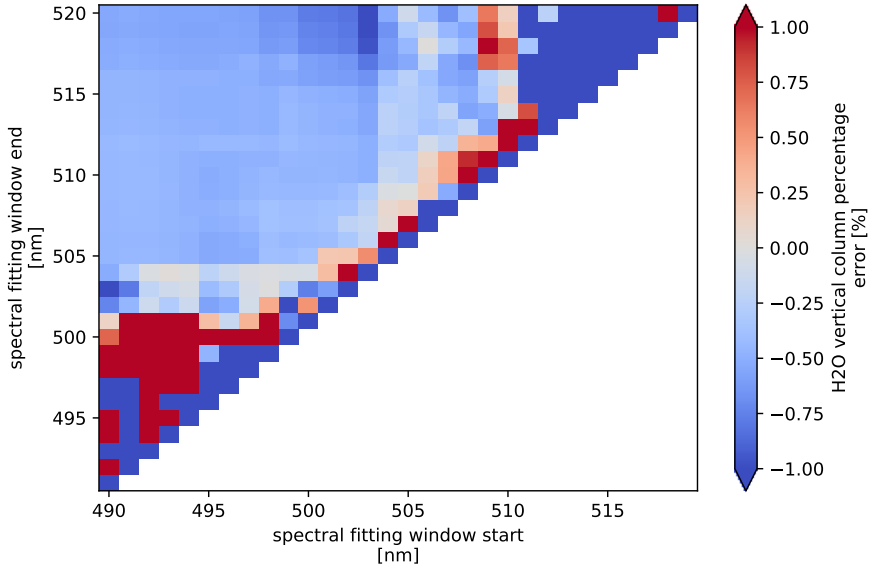
Figure 45: 437-450 nm



490-520 nm (B, Figure 46) The first interesting spectral range that intuitively provides a good signal is from 490-520 nm where the OD typicly exceeds 1e-3. Moreover, this region is also covered by both spectrometer 1 and 2, which allows a generic fitting window since the majority of PGN instruments are 1S.

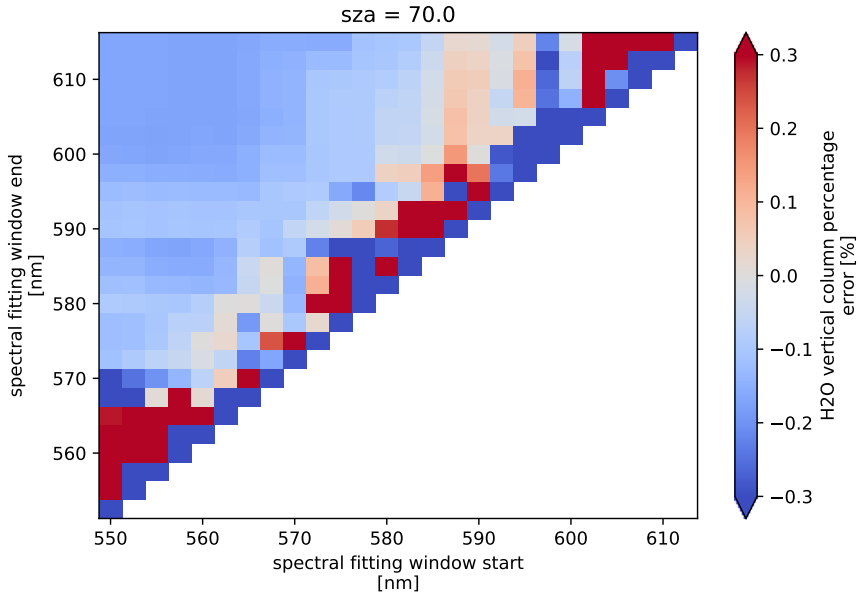
In the spectral range of 490-520 nm O_3, NO_2, O_2O_2, and H_2O are dominant and used for SIMPLE simulations in order to identify the best fitting window. The simulation suggest a short window from 492-503.75 nm. Although there is hardly any temperature dependence of the H_2O OD's, the simulation also reveals an intrinsic bias exceeding 1% error in case of wrong effective temperatures and smoothing polynomials other than 2.

Figure 46: 490-520 nm



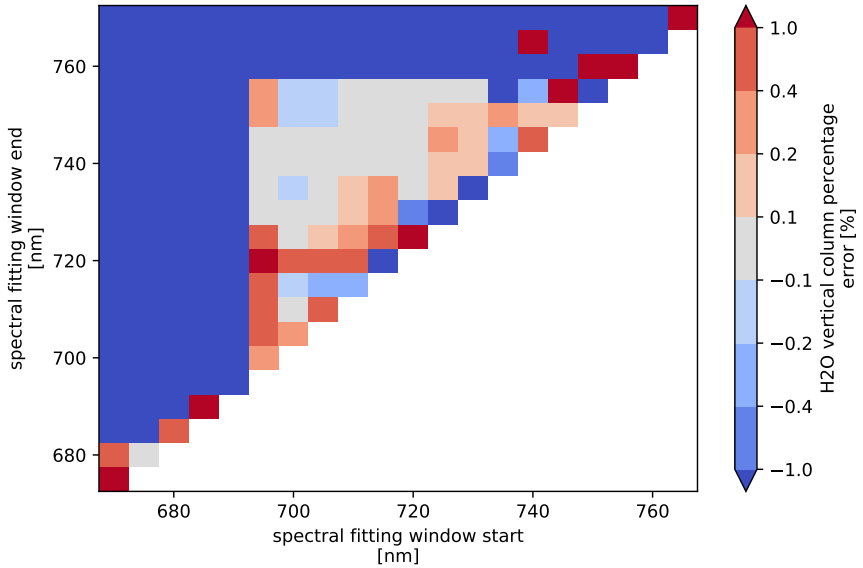
550-615 nm (C, Figure 47) This fitting region shows two strong H_2O peaks, where in particular the first peak between 560-580 nm interferes with the strong O_2O_2 band. The simulation reveals generally two separate fitting windows for each H_2O peak. However, the smallest error is given for 556-567 nm, again with a second order smoothing polynomial.

Figure 47: 550-615 nm



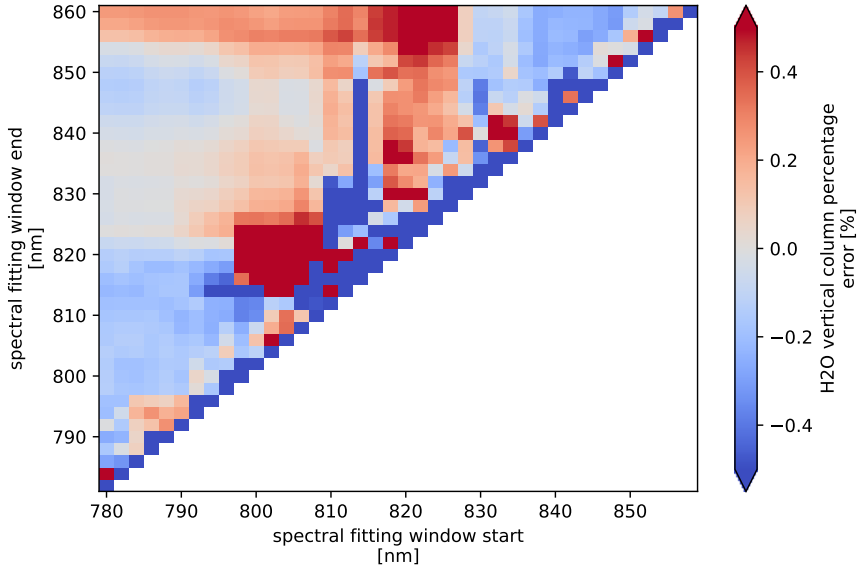
680-770 nm (D, Figure 48) This region is characterized by mainly H_2O and O_3, although a strong O_2 peak up to OD 1e-1 is present from 680 to 700 nm which has to be considered. The OD of H_2O is generally already well above 1e-2, with the larger peak exceeding already 1e-1. The simulations would suggest a well defined region covering this peak between 714-744 nm, which is dominated solely by H_2O and O_3.

Figure 48: 680-770 nm



780-860 nm (E, Figure 49) This region is characterized by only H_2O and O_3 absorption, where OD typically is already well above 1e-1. The simulations highlight a stable region from 784-835 nm.

Figure 49: 780-860 nm



3.12.3 Calibration

In a first test for 'B', the MLE is selected as the calibration technique but has some restrictions. First, the MLE uncertainty is related to the H_2O column variations for the lowest percentile along different AMF's, similar to NO_2, HCHO and SO_2. Second, the MLE can also be affected by the OD method used in the retrieval, since this limits the AMF to be used.

As it will be shown later in the internal validation for H_2O (section 4.7), the MLE can be applicable for the fitting setup 'B'. However, this is not the approach that is envisioned. The ultimate goal is to make use of the AXC approach, as presented for O_3 (Sec.3.5.2). The remaining spectral ranges, where H_2O has OD's well above 1e-2, strong temperature dependencies and highly structured cross-sections, would open the possibility to retrieve H_2O temperatures as well. Furthermore, the AXC calibration approach is applicable on one single day already.

3.12.4 Example applications using a synthetic reference spectra

Figure 50 shows an example time series of total column H_2O for Pandora 138 S1 (left figure panel) and S2 (right figure panel) in black (gray shading reflects the total uncertainty). Data have been taken in Rome, Sapienza university, in early September 2020 and are retrieved using fitting window (and setup) 'B'. For reference, the operational Aeronet H_2O product is depicted in red, which highlights an incredible good agreement between the two data sources, both for S1 and S2.

Two more examples for fitting window 'B' and Pandora S1 at Davos (Pandora 120) and Juelich (Pandora 30) are illustrated in Figure 51. Daily closups are presented in the bottom plots of Figures 52 and 53, respectively. The Pandora data visually highlight a strong correlation in the daily pattern, compared to the co-located Aeronet sites. The Aeronet data are well within the reported total uncertainty. It has to be noted, that the rather large total uncertainty results mainly from the MLE calibration approach, and the climatological H_2O temperature uncertainty.

3.12.5 Example applications using the extraterrestrial reference spectra

Fitting window 'B' is also tested with the extraterrestrial Kurucz reference. Although the synthetic reference is prefered over this approach, a major advantage is the 'out-of-the-box' H_2O data product since no calibration is needed. A first test shows promising fitting results down to 3e-3 fitting wrms. Graphics in the top row of Figures 52 and 53 present daily examples for Davos and Juelich, compared to Aeronet. Pandora data show a very good aggreement in the daily pattern, comparable to the data product using the syntethic reference (bottom row graphics). A quantification of differences is presented in the validation section.
Due to the enhanced temperature dependency of bands C to E, effective temperature fitting needs to be activated for retrievals of theses bands, connected with proper calibration. This task hasn't been performed yet. This is why no field data can be shown for these bands at that point.

3.12.6 Example applications for H_2O temperature fitting

In prior studies, it was not possible to retrieve meaningful slant columns from the visible bands. This was related to the fact that the slit function characterization plays a crucial role, and the calibration process had to be adjusted with respect to omitting doubled lines. In doing so, it was possible to obtain proper slant columns from the strong visible bands. First results imply the possibility to obtain H_2O from all bands, although the final decision for the ideal fitting window is part of ongoing studies.

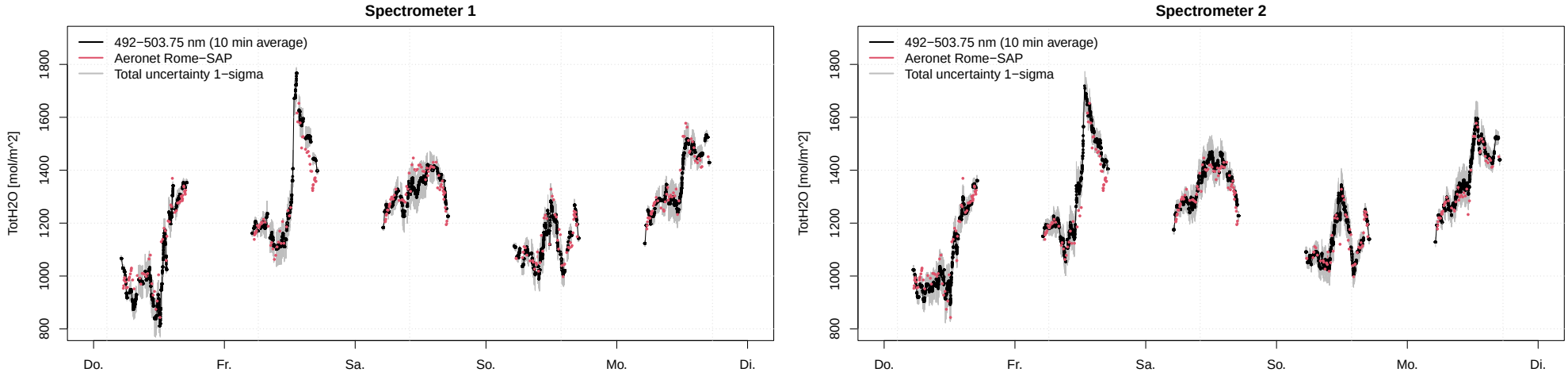
Therefore, an example application for the strongest H_2O band found within the Pandora spectral range is presented. The window 910-930nm, where only O_3 has to be taken into account, has been tested apart from the fitting setups described 3.12.2, since the strongest temperature dependence and H_2O structures are given, which makes the AXC approach applicable for calibration.

In the following, retrievals for the two sites Rome Sapienza and Izana illustrate H_2O direct sun total column amounts and H_2O temperature timeseries.

In addition to Aeronet, the reanalysis satellite composite dataset MERRA-2 offers atmospheric parameters at 42 pressure levels (up to 0.1 hPa $\approx$ 64 km height) at a 3 hour temporal resolution. This dataset does not only provide total columns, but also effective temperatures that can be used for a qualitative comparisons. Moreover, the boundary layer climatology can be used in retrievals without temperature fitting, and is also presented. Since this is not a very meaningful temperature at all, retrieved temperatures for retrievals that explicitly fit the temperature, can strongly differ from the climatology.

Rome As already illustrated in the example application using the synthetic reference spectra, fitting setup 'B' is suitable for spectrometer 1 and spectrometer 2 to retrieve total column amounts, which are comparable to Aeronet amounts. Figure 54 illustrates a 1-year timeseries of the retrieved total columns and the themperature for spectrometer 2. In addition, the effective temperature using the boundary layer climatology is shown. Furthermore, figure 55 presents a close-up day in summer 2020. Visually, the two datasets agree very well in terms of amount and daily variation. With respect the temperature, the retrievals highlight a very low point to point variation, with an increasing temperature over the day. This pattern is also visible in the MERRA-2 dataset, which is negatively biased.

Figure 50: Total H_2O for Pandora 138 at Rome-SAP from 2020-09-03 to 2020-09-08 (black with 1-sigma uncertainty bars) averaged over 10 minutes, and Aeronet data (red). Left figure illustrate spectrometer 1 retrievals, and right figure spectrometer 2 retrievals for fitting setup 'B'.



Izana The same approach as presented for Rome-SAP, has been tested for the new Izana instrument Pandora 209, which has been installed in August 2022, and being fully operational from beginning of November 2022. Figure 56 and 57 present the H_2O retrievals. The total column amounts of Pandora 209 are systematically higher as the Lev 1.5 data product of the corresponding Aeronet instrument, although both instruments agree well in terms of the daily variation (Figure 57). With respect the effective temperature, the MERRA-2 data are matching the retrieved ones.

3.13 Direct sun total O_2

O_2 is the strongest absorber within the Pandora 2S range, with 3 well expressed bands at 630, 690, and 760 nm. O_2 is expected to show less daily variations, and is only pressure and temperature dependent. Therefore direct sun total O_2 can serve as an ideal QC data product similar to O_2O_2 to detect instrumental changes.

A first retrieval setup has been applied in the very narrow spectral range of 759.5 to 775 nm, covering the O_2 peak that exceeds an OD of 1. This is also supported by the simulation (Figure 58). Additional gases that have to be considered are O_3 and H_2O, although H_2O features an OD below 1e-4 for this fitting window. For this first test, a smoothing polynomial of 1 is used.

Figure 59 illustrates direct sun total O_2 columns from spectrometer 2 data of Pandora 138, located at Rome Sapienza.

In a next step, it will be tested how well O_2 can be retrieved from the three individual bands, in support with SIMPLE simulations to optimize fitting windows.

3.14 Sky data surface concentration and tropospheric NO_2 columns

The sky algorithm exploits measurements from 5 pointing zenith angles (0, 60, 75, 88, 89 deg) at a fixed azimuth direction. The measurement sequence is organized in a "V" shape by starting from 0 → 60 → ... → 88 → 89 → 88 → ... → 60 → 0. Since the algorithm is fully parameterized, no elaborate radiative transfer calculations are needed and hence real time data delivery is still given. Currently surface concentration NO_2 and tropospheric columns of NO_2 are extracted from this algorithm. Details about the algorithm can be found in *Cede* [6] and the official publication is in preparation by Elena Spinei Lind et. al.

For PGN data version v1.8, the same quantities for HCHO as well as tropo-

Figure 51: Total H_2O for Pandora 120 at Davos and Pandora 30 at Juelich (black), with Aeronet data in red.

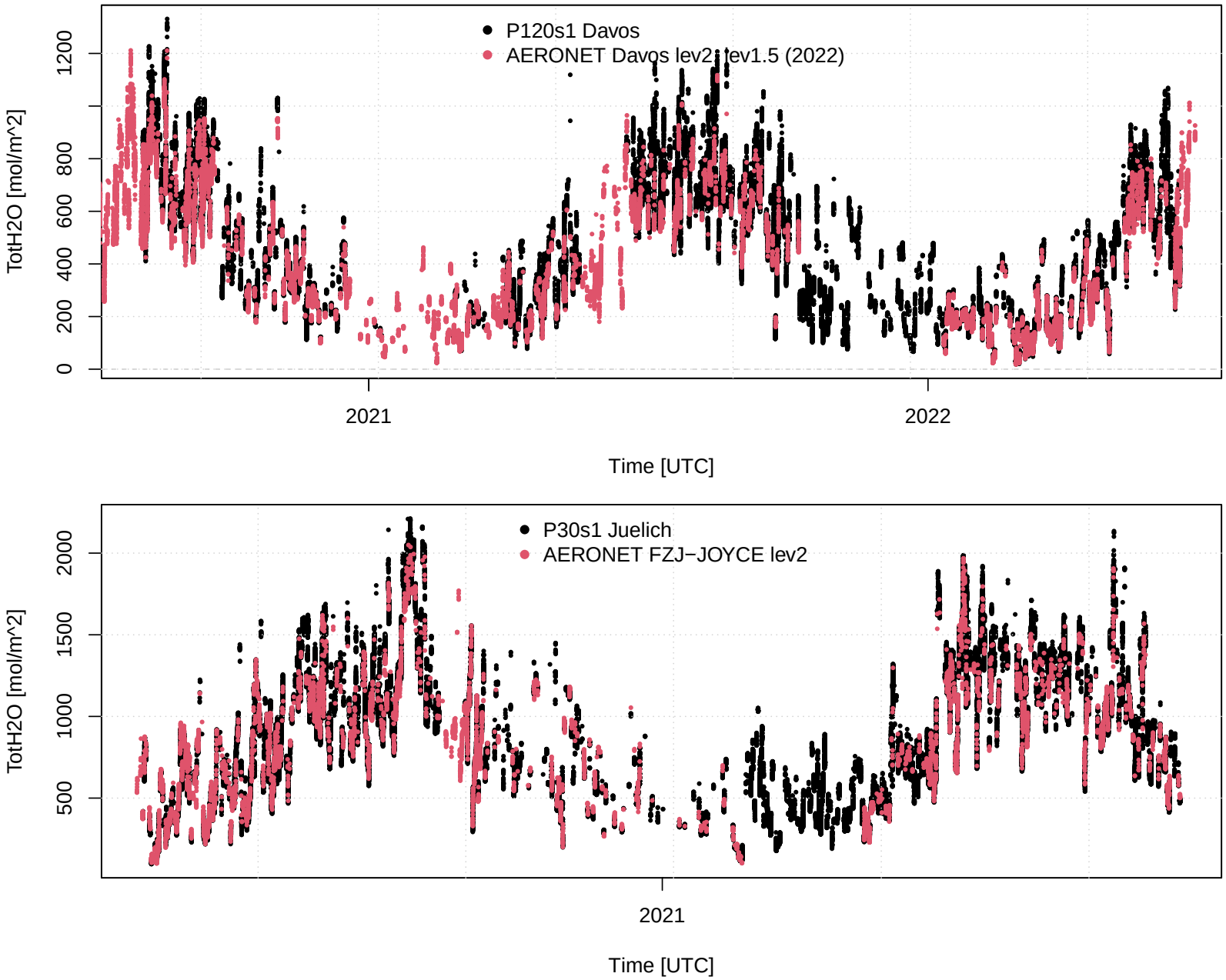


Figure 52: Total H_2O for Pandora 120 at Davos for selected days (columns) using the extraterrestrial reference (top row), and the synthetic reference (bottom row).

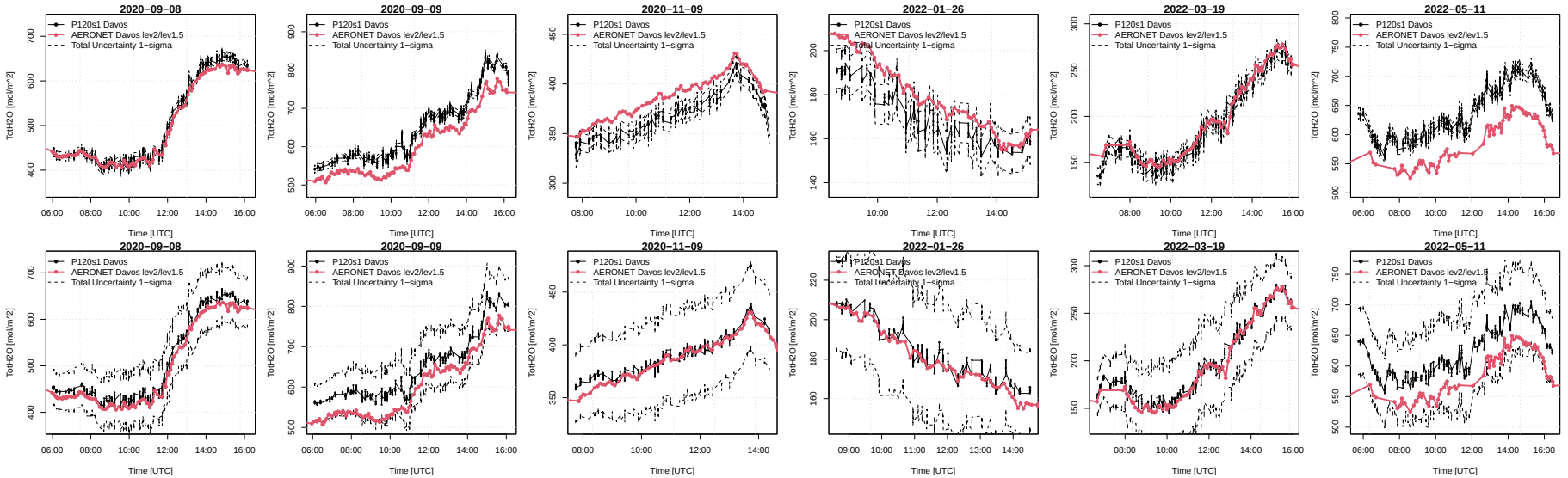


Figure 53: Total H_2O for Pandora 30 at Juelich for selected days (columns) using the extraterrestrial reference (top row), and the synthetic reference (bottom row).

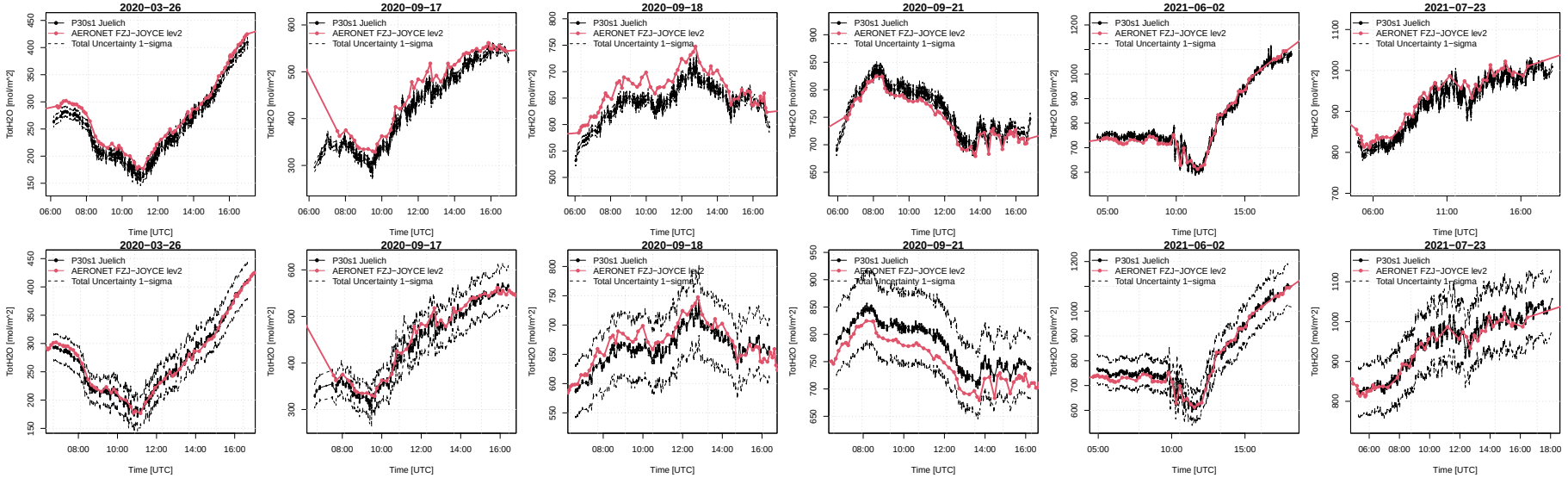


Figure 54: Total column amount of H_2O (top plot) and H_2O temperature (bottom plot) for Pandora 138 spectrometer 2 at Rome-SAP. The grey vertical bars highlight a closeup day.

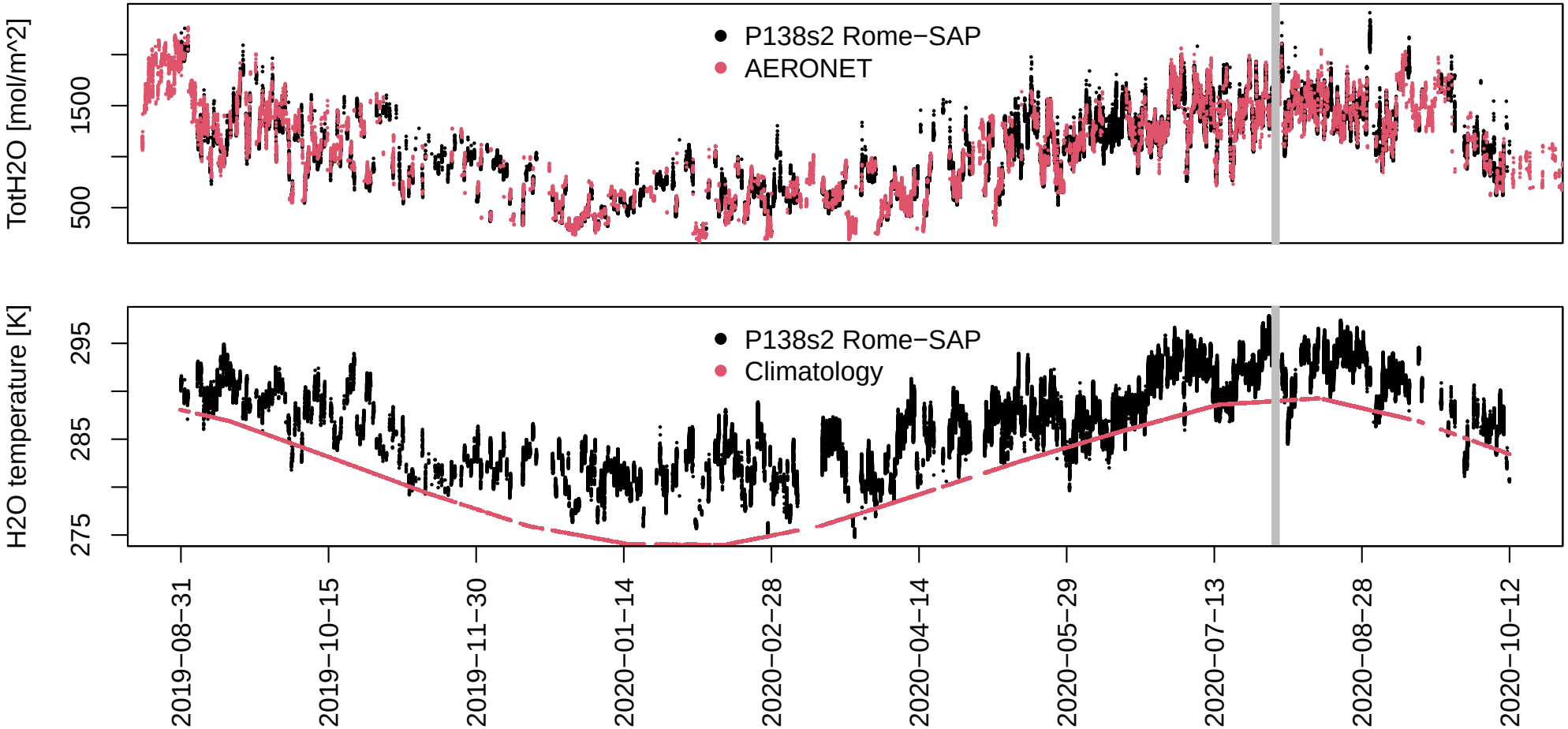
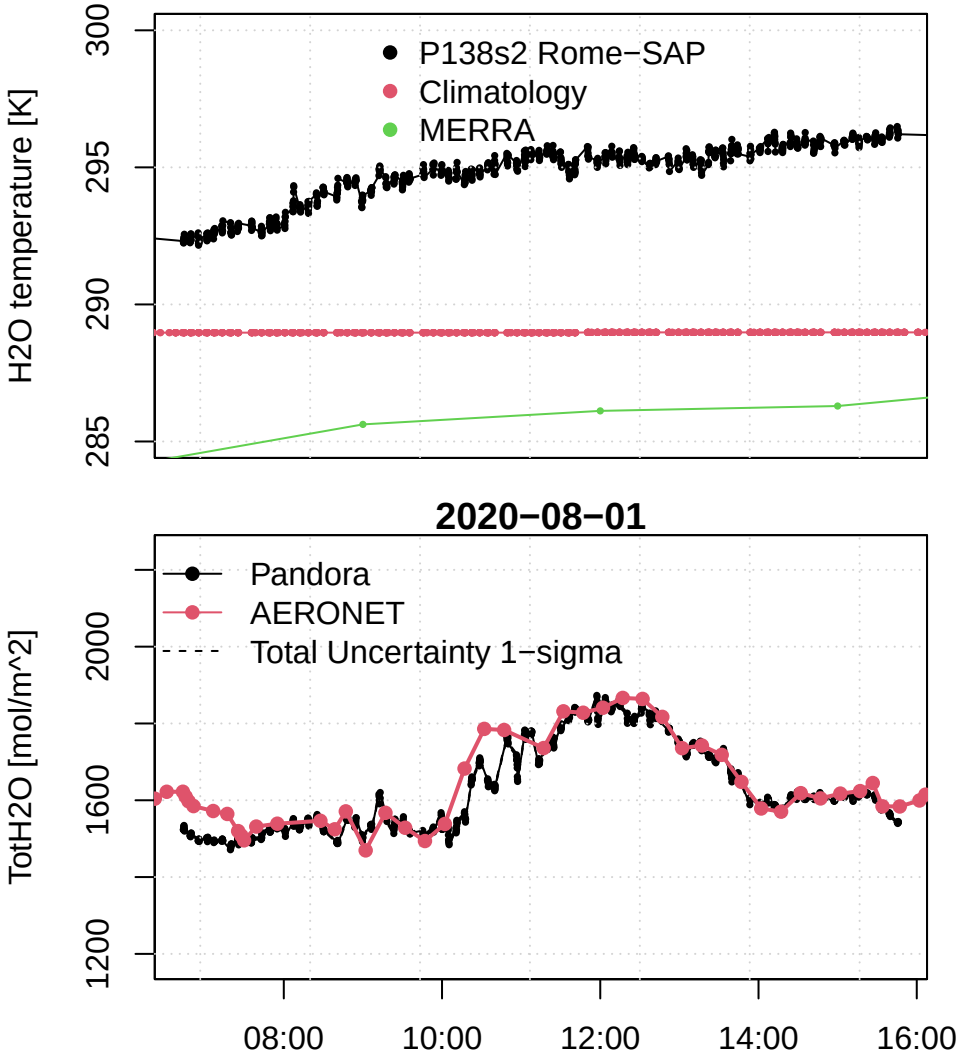


Figure 55: Total column amount of H_2O (bottom plot) and H_2O temperature (top plot) for Pandora 138 spectrometer 2 at Rome-SAP



spheric profiles for both NO_2 and HCHO are planned.

3.14.1 Tropospheric NO_2 columns

The tropospheric NO_2 column product exploits the 60 and 75 deg pointing zenith angle and the zenith measurement as (background) reference. The best guess AMF is estimated from a geometrical approach [30, e.g.] however in a modified way.

A dedicated validation campaign was carried out between September 2016 and August 2017 in Innsbruck, Austria. During that time period two Pandoras where placed in a way that one instrument was sited in the valley center and the other instrument in a horizontal (vertical) distance of 4.5 (1.6) km at the mountain top of "Hafelekar" (HAF) (compare Figure 60). Both Pandoras measured both direct sun and sky measurements. With this special setup, by taking the difference between the total column amounts of the mountain Pandora minus the valley Pandora, the column amount of the valley atmosphere can be calculated very precisely. The valley atmosphere is representative for the tropospheric columns in cases where the boundary layer stays below the crest height. This is the case for Innsbruck beginning late autumn, winter and early spring. Only this data is used for this validation study.

Figure 60: Orographic map of Innsbruck, Austria, with its mountainous surrounding. The Pandora located in the valley center is marked with IBK (for Innsbruck) and the one on the mountain site with HAF (for Hafelekar). The pointing azimuth angle of the valley Pandora is towards East and illustrated by the blue lines.

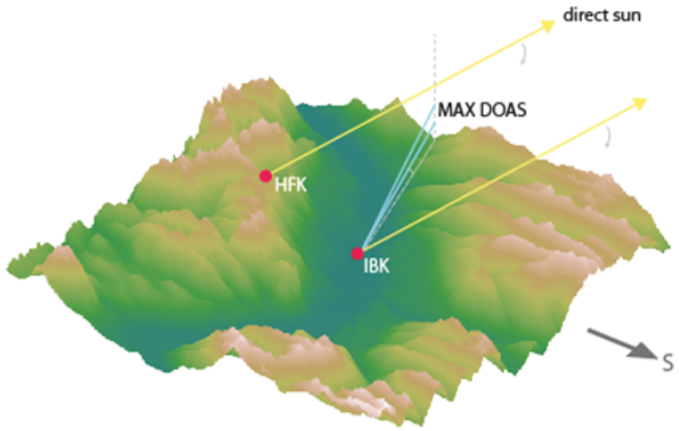


Figure 56: Total column amount of H_2O (top plot) and H_2O temperature (bottom plot) for Pandora 209 spectrometer 2 at Izana. The grey vertical bars highlight a closeup day.

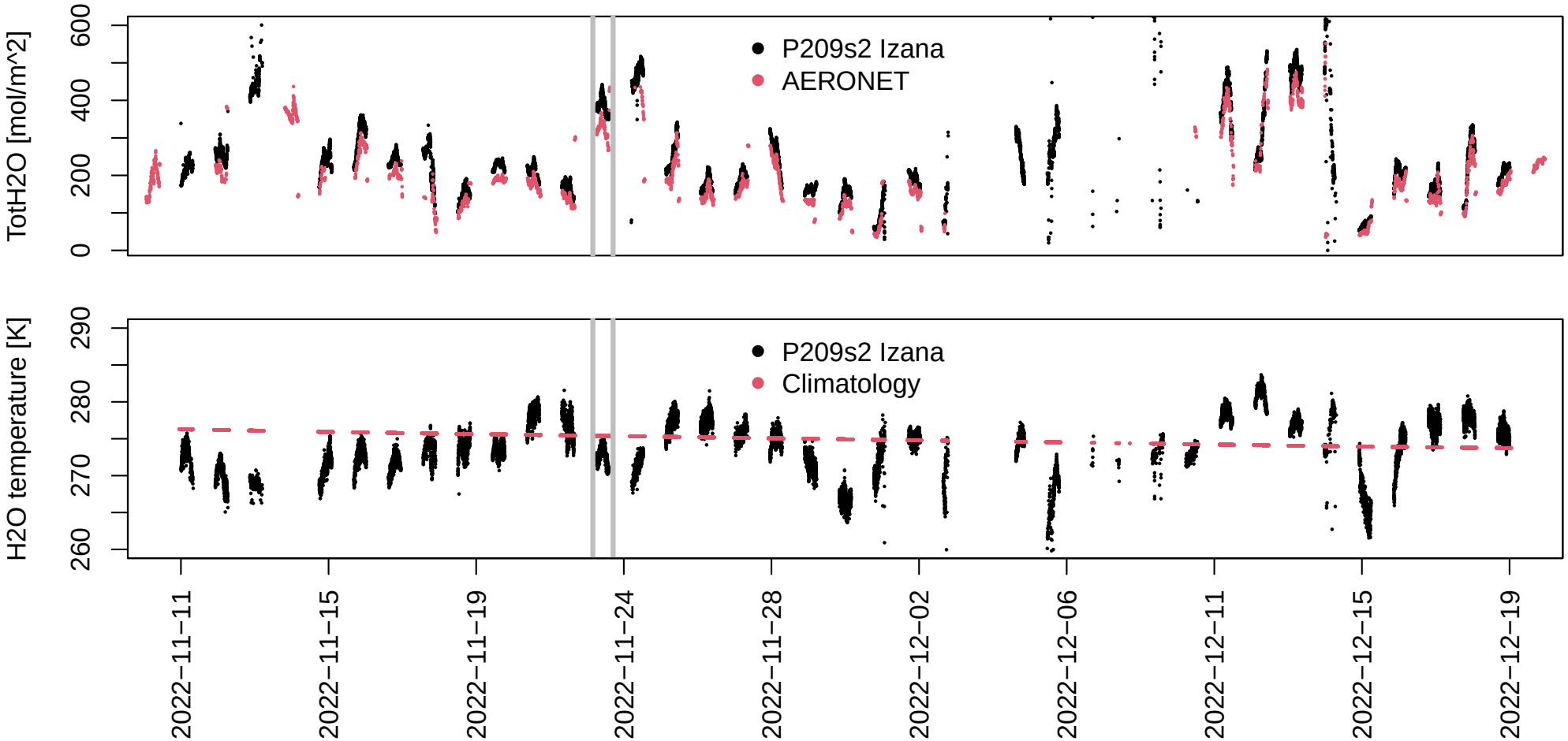


Figure 57: Total column amount of H_2O (bottom plot) and H_2O temperature (top plot) for Pandora 209 spectrometer 2 at Izana

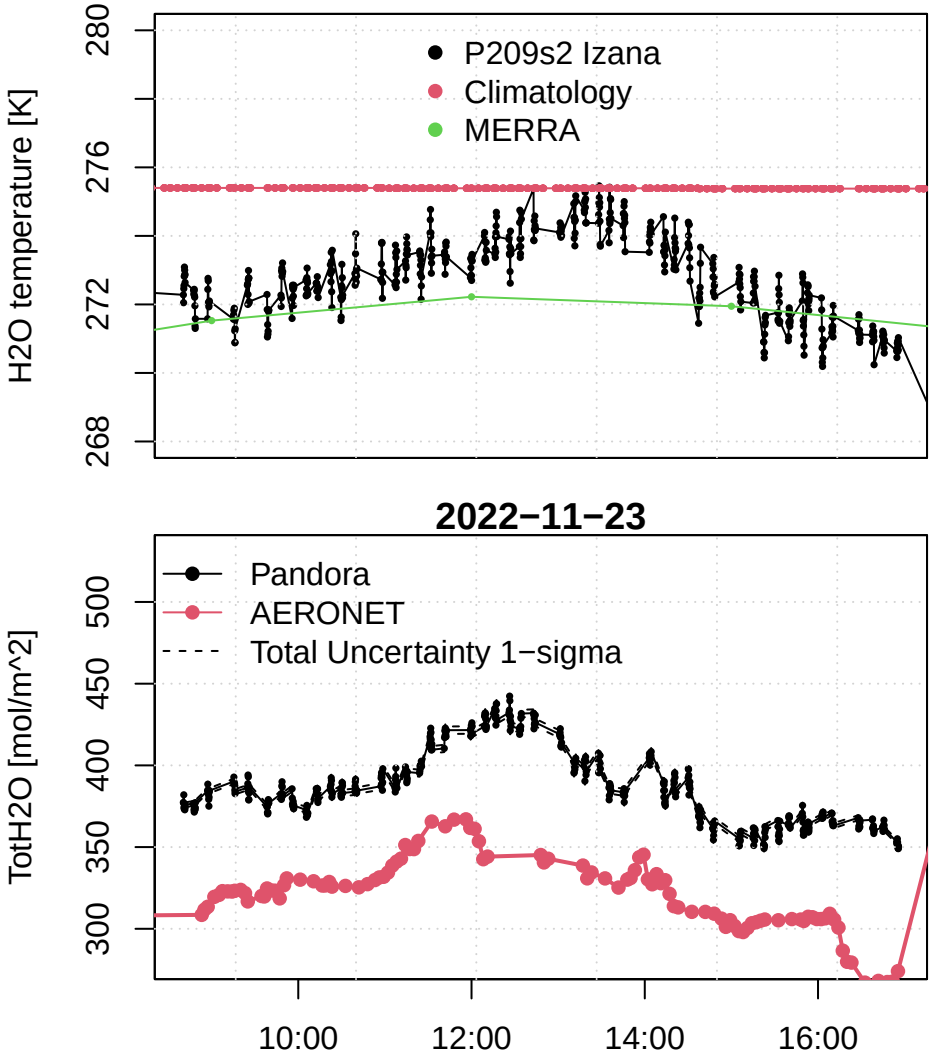


Figure 58: Simulated O_2 vertical column errors for SZA=70 as a function of start and end wavelength.

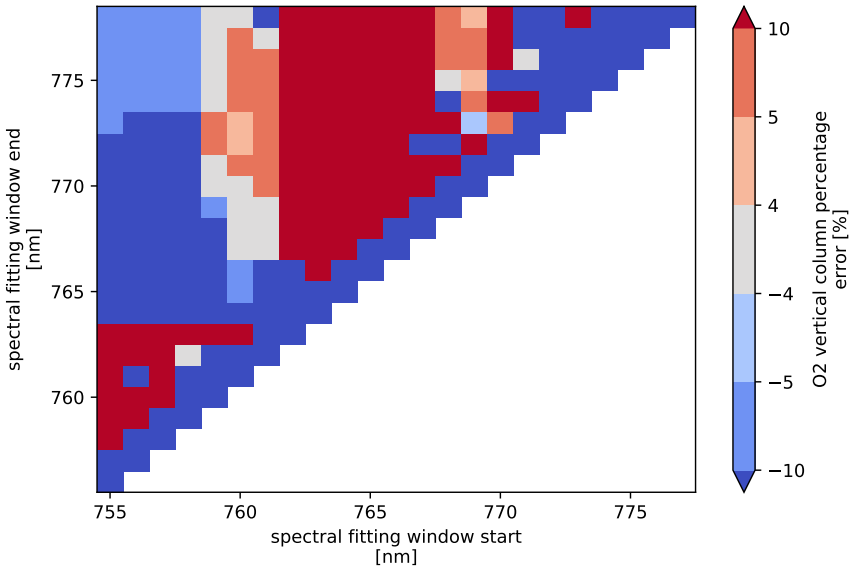
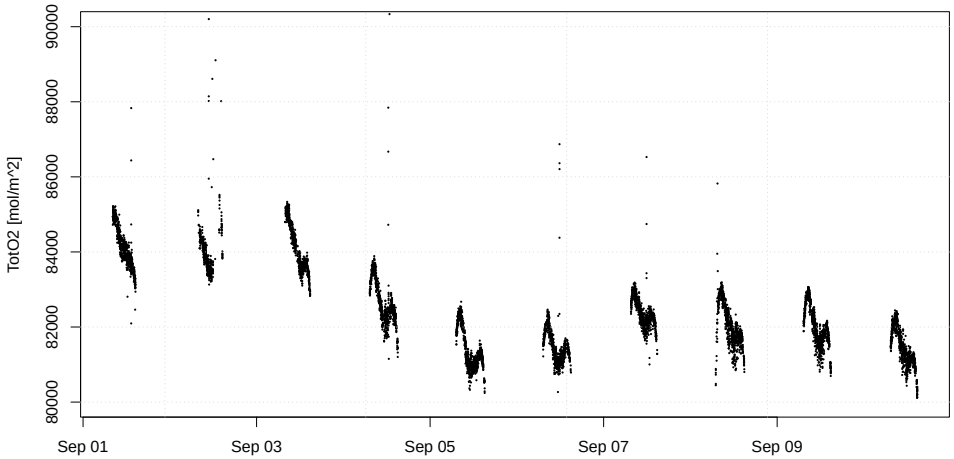
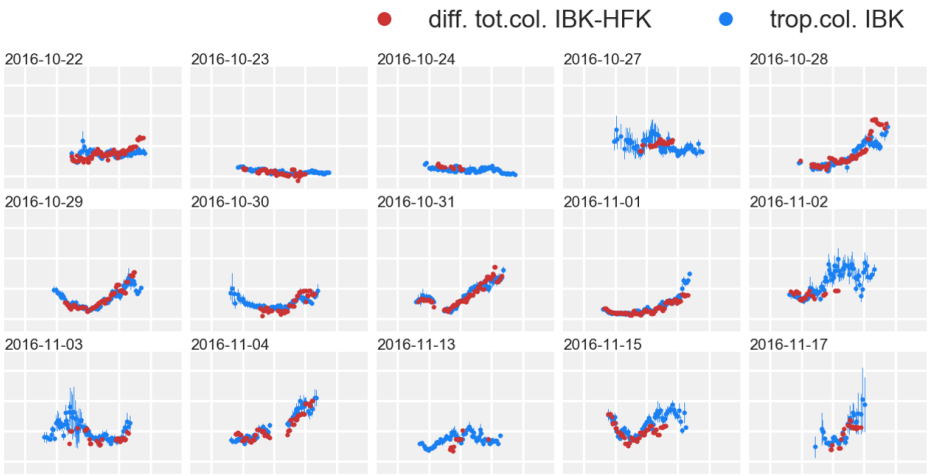


Figure 59: Direct sun total O_2 column for P138s2 at Rome Sapienza



A selected time period of the comparison is shown in Figure 61 where red dots mark tropospheric columns from the sky algorithm from the valley Pandora and blue dots the counterpart from the difference of the mountain and valley total columns. The agreement is exceptionally good with differences mainly in the afternoon where the observational azimuth directions differ most (compare Figure 60). Evaluating the entire period confirms the good agreement with a very small median difference of 0.03 DU and a 1σ spread of 0.13 DU.

Figure 61: Comparison of tropospheric NO_2 columns retrieved from two different methods. Blue dots show the retrievals based on the sky scans measured by the Pandora in the valley. The difference of total columns from the Pandora located at the mountain site and the valley is gives as red dots. The y-axis spans the range from -0.1 to 1.7 DU and the x-axis shows the time from 5 to 20 o'clock local time.



Owing to the very good validation results the tropospheric NO_2 product from sky data is ready to be included as standard PGN product. Further, the applicability for the new E-MLE calibration (see section 3.3)) is confirmed.

3.14.2 Tropospheric NO_2 columns from spectrometer 2

As for direct sun measurements (see 3.3.1), the advantage of including tropospheric measurements of S2 would be to have a time synchronous measurement when there is a measurement with S1 and further, the S2 measures NO_2 when the S1 does a U340 measurement. Due to the much lower signal of sky measurements compared to direct sun measurements, no diffuser is needed and the measurements are either taken with OPEN or a U340 filter with the S1. Consequently the S2 measurements are both taken with OPEN, just on different positions of the filterwheel. The huge benefit of the sky retrieval is that the reference is taken for each measurement set, so we do not have the problems with the individual filters as we have for the direct sun products as shown in section 3.14.2.
S2 measurements, which are taken in parallel with the S1 measurements, agree very good as shown in table 4 and figure 62. There is a systematic bias with the S2 having slightly higher tropospheric NO_2 columns, which could be related to the broader slit function of the S2, but the bias is very small as shown in table 4. This means we also can trust the S2 measurements which are taken when S1 is doing U340 measurements.

Table 4: Comparison of tropospheric NO_2 difference between S1 and S2 of P117 and P138. 416 measurements from 1st August 2020 till 15th September 2020 have been used for P138 and 368 measurements for P117. The time lag between the measurements does not exceed one second. Units are given in $mmol/m^2$.

	P138s1 vs. P138s2	**P117s1 vs. P117s2**
median difference	1.61e-3	1.18e-3
mean difference	1.58e-3	1.87e-3
σ	1.89e-3	2.28e-3

Figure 62: The tropospheric NO_2 difference between P138s1 and P138s2. 416 measurements from 1st August 2020 till 15th September 2020 are shown and the time lag between the measurements does not exceed one second.

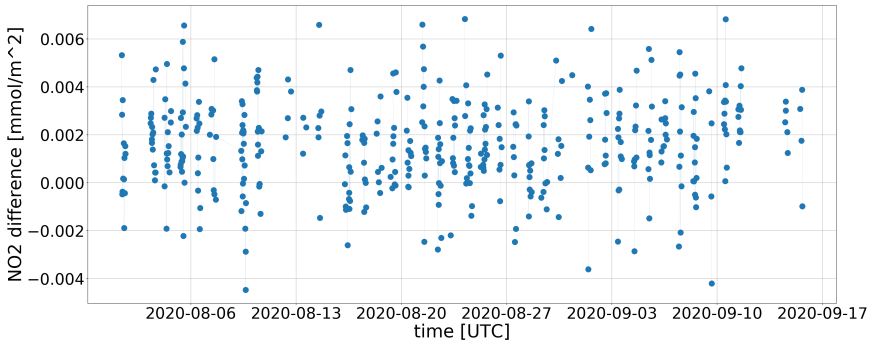


Figure 63: The tropospheric NO_2 of one day of Pandora 117. The red and green dots are measured within the same elevation scan and the blue data points of S2 can nicely fill up the gaps.

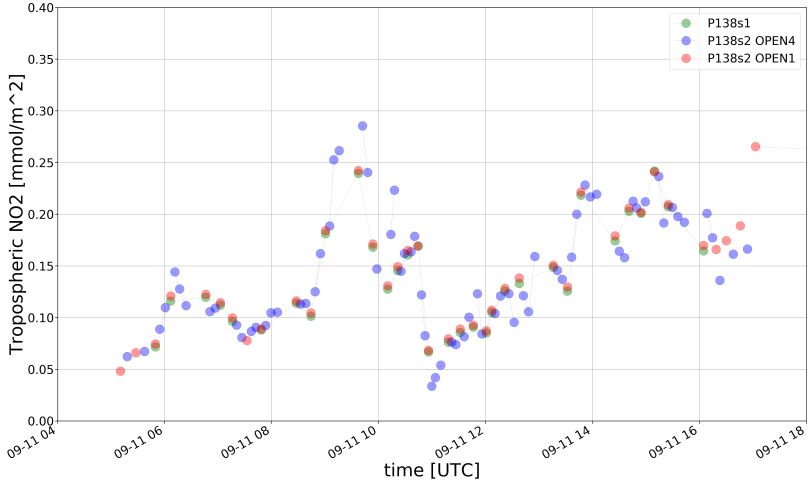


Table 5 shows a comparison of two different instruments. Both, S1 and S2, show differences in the same order of magnitude. The uncertainty of the S2 measurements is often better due to the higher sensitivity of the S2. This is not always the case since it also depends on the filter setup and the used attenuation filter in the field. However, for all instruments built or upgraded since 2019, having the ND filter setup with finer steps, the uncertainty is expected to be better for the S2. This means the tropospheric NO_2 of S2 can be included as additional data product in the PGN.

Table 5: Comparison of tropospheric NO_2 difference between P117 and P138 for 1st August 2020 till 15th September 2020 for both spectrometers. For every parameter there are two numbers shown which corresponds to the time lag af 30s and 60s between different measurements. Units are $mmol/m^2$

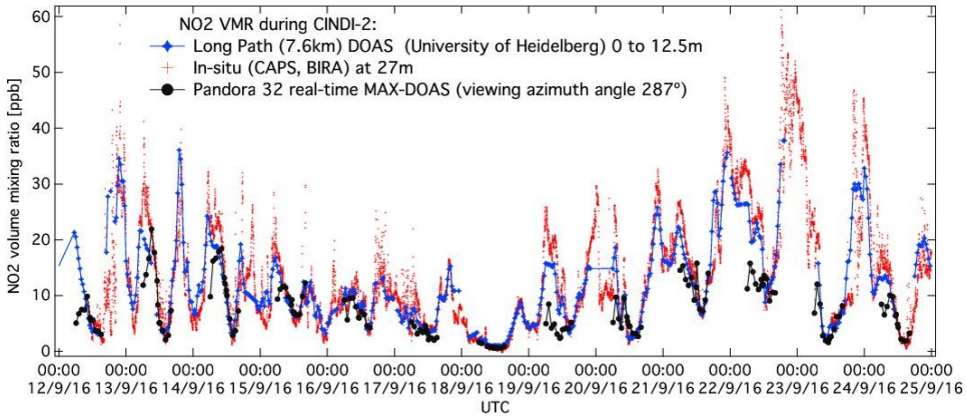
	P117s2 vs. P138s2: 30s, 60s	**P117s1 vs. P138s1: 30s,60s**
median difference	1.03e-3, 9.78e-4	1.34e-4, 3.08e-4
mean difference	9.79e-4, 1.09e-3	5.03e-4, 5.12e-4
σ	4.7e-3, 5.93e-3	3.89e-3, 4.63e-3
N	535, 1025	65, 102

3.14.3 Surface concentration NO_2

The NO_2 surface concentration product makes use of the two uppermost pointing zenith angles, usually 88 and 89 deg and the zenith measurement as (background) reference. Since the retrieved slant columns are extrapolated to 90 deg, we indeed get surface rather than near surface concentrations.

This data product underwent already a number of validation studies carried out by Elena Spinei Lind, which are outlined in detail in Lind et. al, 2021 (in preparation). One example from the CINDI-1 instrument intercomparison campaign [39] is shown in figure 64, which correlates Pandora data (black), long path DOAS (operated by University of Heidelberg) retrievals (blue) and CAPS in-situ measurements (operated by the Royal Belgian Institute for Space Aeronomy (BIRA)).

Figure 64: Surface concentration comparison (CINDI-1 campaign, Netherlands) of NO_2 for three different measurement methods. Pandora data are shown in black, long path DOAS data in blue and CAPS in situ data in red.



The validation results suggest this data product to be a valuable addition of the official PGN data products.

3.14.4 Surface concentration NO_2 of spectrometer 2

Since surface concentration measurements are retrieved out of the same measurement set as tropospheric data, the same is valid as presented in section 3.14.2 for tropospheric NO_2. The results are outlined in this section. The standard deviation between time synchronous measurements are shown in table 6 and are in the 2e-8mol/m^2 range for P117 and below 1e-8mol/m^2 for P138.

Table 6: Comparison of tropospheric NO_2 difference between S1 and S2 of P117 and P138. 416 measurements from 1st August 2020 till 15th September 2020 have been used for P138 and 368 measurements for P117. The time lag between the measurements does not exceed one second. Units are given in mol/m^3.

	P138s1 vs. P138s2	**P117s1 vs. P117s2**
median difference	5.1e-9	6.5e-9
mean difference	5.7e-9	9.4e-9
σ	6.3e-9	1.9e-8

Table 7: Comparison of surface concentration NO_2 difference between P117 and P138 for 1st August 2020 till 15th September 2020 for both spectrometers. For every parameter there are two numbers shown which corresponds to the time lag af 30s and 60s between different measurements. Units are mol/m^3

	P117s2 vs. P138s2: 30s, 60s	**P117s1 vs. P138s1: 30s,60s**
median difference	5.7e-9, 5.7e-9	8.0e-9, 7.4e-9
mean difference	6.0e-9, 3.6e-9	9.3e-9, 8.2e-9
σ	4.4e-8, 5.7e-8	1.0e-8, 9.6e-9
N	535, 1025	65, 102

Figure 65: The surface concentration NO_2 difference between P138s1 and P138s2. Measurements from 1st August 2020 till 15th September 2020 are shown and the time lag between the measurements does not exceed one second. As for the tropospheric NO_2 columns, the S2 shows systematically higher values.

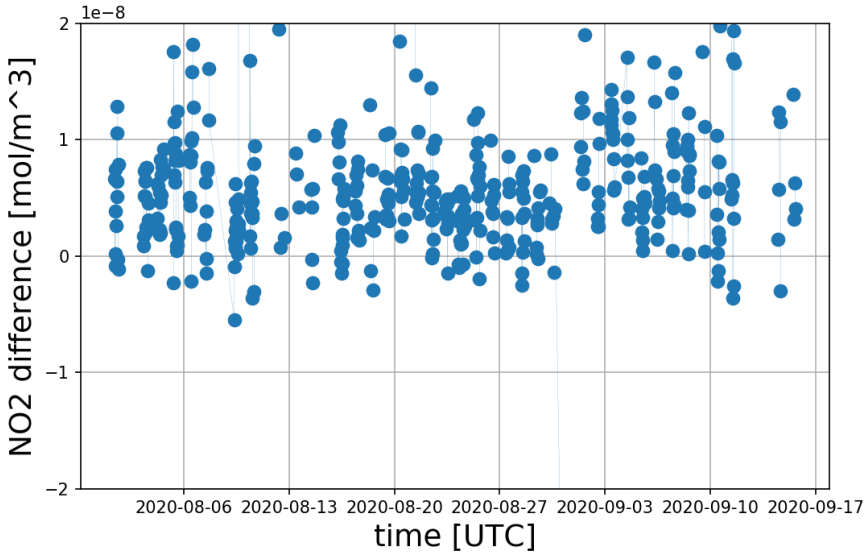
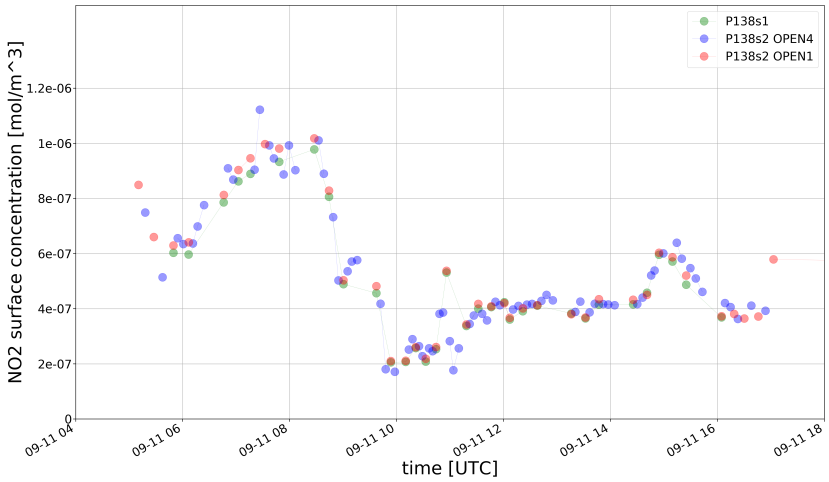


Figure 66: The surface concentration NO_2 of one day of Pandora 117. The red and green dots are measured within the same elevation scan and the blue data points of S2 can nicely fill up the gaps. OPEN1 and OPEN4 refers to two different OPEN filterwheel positions, but physically the same is measured.



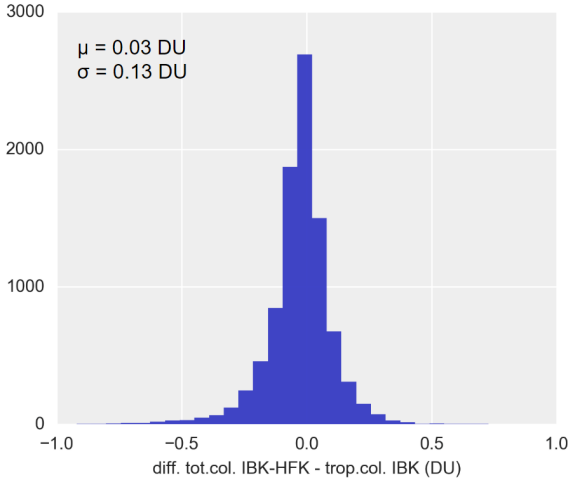
3.15 Sky data surface concentration and tropospheric HCHO columns

The algorithm for MAX-DOAS HCHO products (surface concentration and tropospheric columns) is already quite matured. It could clearly be shown during the CINDI-2 campaign that the Pandora is sensitive enough to retrieve HCHO form sky data [49], and the surface concentration of HCHO compares well to reference data sets (Spinei, in preparation).

The before mentioned validation campaign for NO_2 can unfortunately not be applied to the validation of tropospheric HCHO. The involved Pandoras are prone to the "Delrin issue" and hence cannot deliver proper direct sun HCHO data.

Due to the need of tropospheric HCHO in order to make an E-MLE calibration approach possible for total HCHO, the readiness of both data products is closely connected.

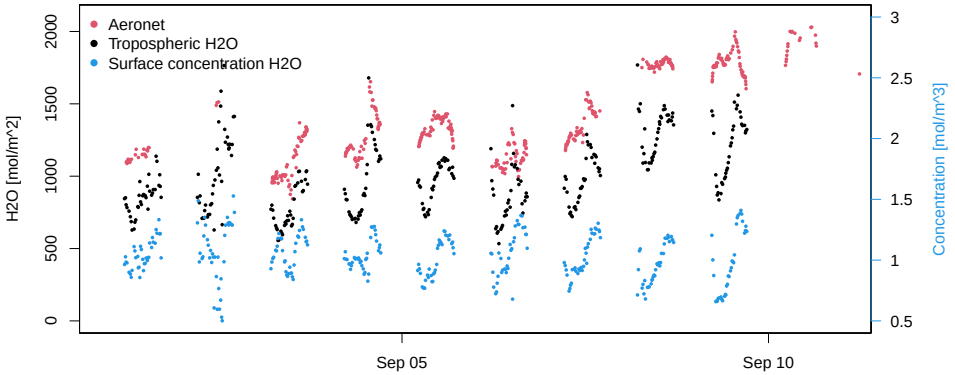
Figure 67: The difference of NO_2 total columns measured by a Pandora in the valley and on the mountain top is compared to tropospheric NO_2 retrieved from a sky algorithm for the valley Pandora.



3.16 Sky data surface concentration and tropospheric H_2Ov columns

H2O from sky data is retrieved from the same fitting setup as used for NO_2 data (Sec.3.14). A timeseries for Pandora 138s1 at Rome Sapienza is illustrated in Figure 68.

Figure 68: Tropospheric (black) and surface concentration (blue) H_2O at Rome Sapienza. Aeronet H_2O values are shown in red.



Surface concentrations and tropospheric columns agree in terms of their diurnal variation. This example also shows an interesting advection pattern: while the average surface concentrations stay approximately constant, the tropospheric columns increase over the days. Further, the intrinsic difference[7] between Aeronet total column data and tropospheric columns also increase. These observations point towards advection of H_2O in higher layers.

Since it was found from direct sun H_2O retrievals that H_2O can be well retrieved using the 500nm H_2O peak, an adjusted sky fitting window for tropospheric H_2O should be tested.

3.17 Sky data profiles

In order to meet the demand on real time data delivery, the profile algorithm implemented in BSSv1.8 follows a parameterized approach again. It is based on the already introduced sky algorithm and it delivers gas concentrations for as many layers as viewing zenith angles measured. This algorithm has been developed by Elena Spinei Lind in collaboration LuftBlick.

An example is illustrated in Figure 69, where one can see profiles of NO_2 (top panel), H_2O (middle panel) and HCHO (bottom panel) in units of mol m^{-2} km^{-1} for some selected days in Rome (Sapienza university, Pandora 117).

[7]The limited vertical sensitivity of MAX observations (up to about 3-4 km) basically only captures the boundary layer and parts of the free troposphere and by this does not "see" a significant portion of the vertical water content. Hence a discrepancy between integrated MAX and DS columns is intrinsic.

3.18 Spectral AOD

When deriving spectral aerosol optical depth (AOD) one faces very different challenges than for trace gas retrievals. While things line wavelength accuracy, exact knowledge of the slit function or high signal to noise are of minor influence, issues such as radiometric stability and the shape of the field of view (FOV) are extremely important. A study on the retrieval of spectral AOD with Pandora is given as an appendix B. It shows that a significant amount of tasks need to be done in order to make spectral AOD an operational Pandora data product.

3.18.1 Stage one improvements

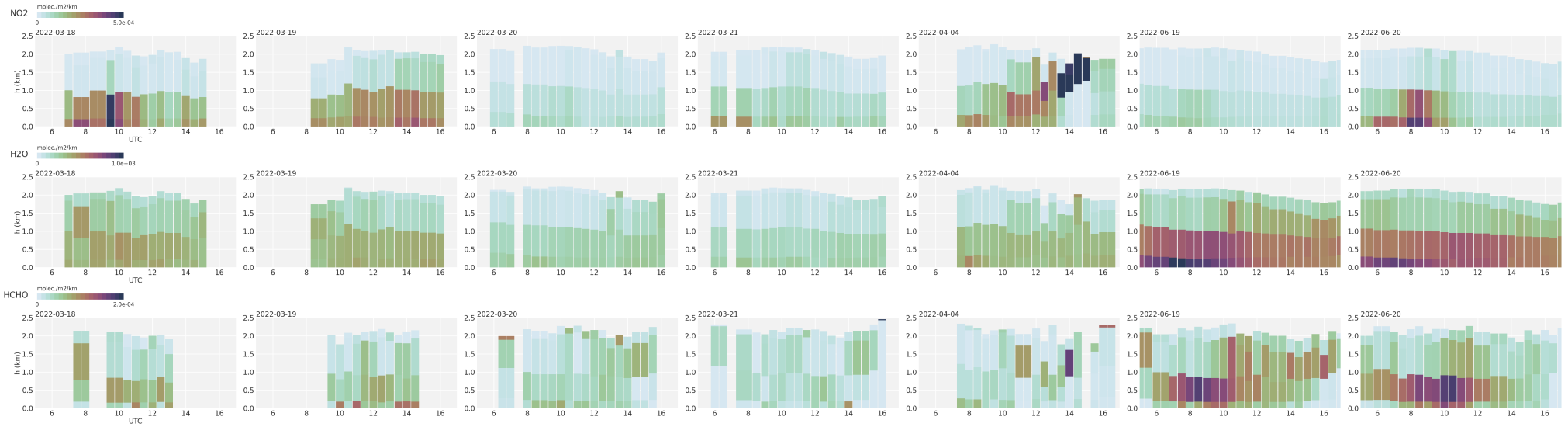
The first group of the suggested tasks (5 tasks, titled "stage one improvements" in the appendix) could be implemented but one. Those are:

L1 data are corrected for spectral stray light
As part of the improvements within the BSS v1.8, the spectral stray light correction based on the "matrix-method" can now be applied to Pandora L1 data (refer to section 3.2.2). In contrary to the O_3 retrieval - where in particular the near field stray light has a large impact - far field stray light correction is crucial for AOD retrievals too. The consideration of outstanding features in the far field stray light (like the "ghost" feature visible in Figure 1) leads to further improvement of the stray light correction quality.

AXC is used to get accurate total column O_3
The new O_3 products (introduced in section 3.5.2) are able to provide accurate O_3 columns for the AOD retrievals. In particular for sites showing significant tropospheric O_3 pollution, the connected O_3temp change can be considered explicitly (in the SyntRef product). This further reduces uncertainties in the O_3 retrieval.

Each instrument has a fiber guide
The production of the fiber guides is done at SciGlob as part of the Pandora production. Hence all new Pandoras are shipped with fiber guide. All existing Pandoras have been equipped with fiber guides within the framework of this project.

Figure 69: Trace gas profiles of NO_2 (top panel), H_2O (middle panel) and HCHO (bottom panel) in units of mol m^{-2} km^{-1} for some selected days in Rome (Sapienza university, Pandora 117).



The shape of the FOV is improved

The quality of the FOV is primarily dependent on the alignment of the fiber relative to the optical system of the sensor head (e.g. lens) and the manufacturing quality of the fiber front surface (homogeneity and perpendicularity). SciGlob works on means to improve the first point but is dependent on the manufacturer regarding the latter point.

The 5th task of "stage one improvements" - the improvement of the calibration and stability of the calibration constant - could not be completed yet. This is due to the fact that no mobile reference unit has been available so far.

3.18.2 Stage two improvements

Two improvement tasks have been suggested for stage two. The one related to software improvements could be accomplished, however the hardware related task hasn't been realized yet.

A pressure sensor is added to the Pandora

While in the meanwhile the majority of the Pandoras are equipped with a pressure sensor in the sensor head, an external pressure sensor has not yet been added to the system and software.

A water vapor retrieval is developed

A H_2Ov product has been developed for the 505 nm absorption band and is introduced in section 3.12. The selected band has the advantage that it is covered by both spectrometers (UV and VIS) and features a small dependency on the effective temperature.

4 Data product validation

4.1 Validation strategy

We approach the validation of the data products in two ways: first, we do what we label **internal validation**. This validation approach makes use of Pandora data only. Either

- independently calibrated direct sun products of several co-located Pandoras are intercompared (InterComp), or
- direct sun and sky data retrievals of a tropospheric absorber for one Pandora are intercompared (CrossComp).

At a second stage we perform **external validation**, that is comparing the data products to independent datasets.

4.2 Direct moon NO_2 validation

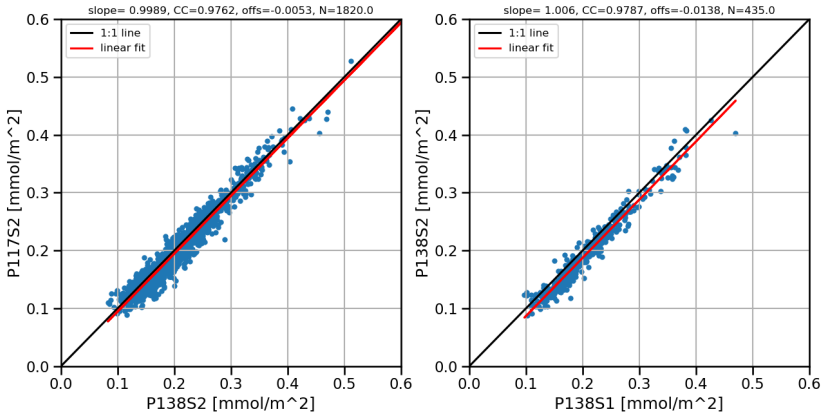
4.2.1 Internal validation

Between end of July 2020 and October 2020, two 2S-Pandoras, Pandora 117 and 138 have been co-located in Rome-SAP which provides an excellent dataset for internal validation of lunar NO_2. In figure 70, the comparsion of these two instruments for the above mentioned timeframe is shown. Within this timeframe, three full moon cycles have been present. With the used filter criteria of the wrms not exceeding 2e-3 and the uncertainty not being higher than 2e-2 mmol / m^2, we had datapoints for the spectrometer 2 for roughly $\pm$ 6 days around full moon and $\pm$ 3 days around full moon for spetrometer 1, which explains also the lower number of measurements available for the comparison of a spectrometer 2 towards a spectrometer 1 in figure 70. For comparing two different instruments the spectrometer 2 has been used due to the lower uncertainty and higher amount of filtered measurements. When averaging the data by three minutes, the agreement between two different instruments is very good with a slope of almost 1, the correlation coefficient above 0.97 and a offset of 0.0053mmol / m^2. When comparing spectrometer 1 towards spectrometer 2 of one instrument, less datapoints are left for the comparison due to the overall lower amount of filtered data. The slope and the correlation coefficiant are very good, whereas the slope is a little bit higher being 0.0138 mmol / m^2. Table 8 shows comparison results for Pandora 117S2 and 138S2. Due to the excellent agreement for 3min averages, we can conclude that spectrometer 2 lunar data can resolve tropospheric NO_2 pollution on a minute basis and longer time averaging is not needed since it would smear away information.

Table 8: Comparison table of direct moon NO_2 data for the VIS spectrometers between Pandora 117s2 and Pandora 138s2 for different averaging times.

avg time	slope	intercept [mmol/m^2]	CC	N
1min	1.0059	-0.0062	0.9772	1035
3min	0.9989	-0.0053	0.9762	1820
5min	1.0151	-0.0087	0.9778	1255
10min	1.0119	-0.0077	0.9794	704
30min	0.9371	0.0064	0.9847	281

Figure 70: 3min averages of direct moon data. On the left, the comparison between two different instruments for the spectrometer 2 is shown. The right side shows the comparison between spectrometer 1 and spectrometer 2 of P138.



4.2.2 External validation

There are not many lunar datasets available from other ground based systems and we have not been aware of other lunar data overlapping with Pandora data. But the data can be compared towards a stratospheric climatology (from OSIRIS measurements, further described in *Brohede et al.* [17]) dependent of the latitude, season and time of the day.

In figure 71 one can see that the daytime difference is higher as the difference during night for Pandora 121 at Izana. The daytime difference is roughly 1e-5mol / m^2 and the nighttime difference between 3e-5 and 8e-5mol / m^2 for the lunar data around full moon on 18th May 2019. Due to NO_2 residuals in the free troposphere, which can be seen by Pandora but which is not part of the stratospheric climatology, a bias between Pandora measurements and the climatology is understood, whereas the higher difference for lunar data is something which will have to be further examined in the future.

A similar comparison can be made for another background location, WrightwoodCA (close to the NASA Jet Propulsion Laboratory's Table Mountain Facility (TMF)). In *Li et al.* [34], a dataset of direct sun and moon NO_2 data from a grating spectrometer, retrieved with the DOAS principle has been compared towards a photochemical model developed by Caltech. There is no direct time overlap with Pandora measurements, but since there is hardly any tropospheric pollution for most of the time, one can draw conlusions by comparing this dataset towards Pandora data of the same season two years later. Figure 72 shows data of this location for 2018 from the JPL measurements and 2020 for Pandora measurements. The daytime difference for Pandora data and the climatology is around 1e-5mol/m^2 - 2mol/m^2 and the nighttime difference ranges from 3e-5mol/m^2 to 8e-5mol/m^2. For the JPL difference towards the climatology we see values between 4e-6mol/m^2 and 2e-5mol/m^2 during day and 2e-5mol/m^2 - 4e-5mol/m^2 during night. So the nighttime difference is higher as the daytime difference for the JPL measurements and more pronounced for the Pandora measurements. Concluding, this might be a general problem for lunar NO_2 measurements and will have to be further examined regarding Pandora data.

Figure 71: Direct sun (5min averages) and direct moon (15min averages) NO_2 for Izana with the stratospheric climatology (half hourly steps) for ± 1 day around full moon. The spectrometers with the better total uncertainties are shown.

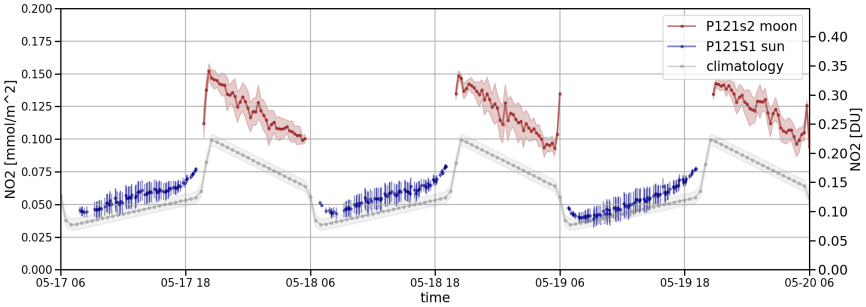
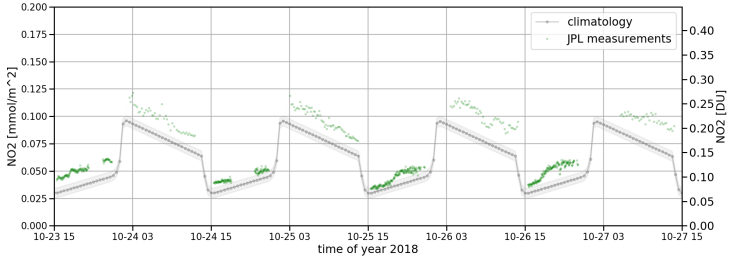
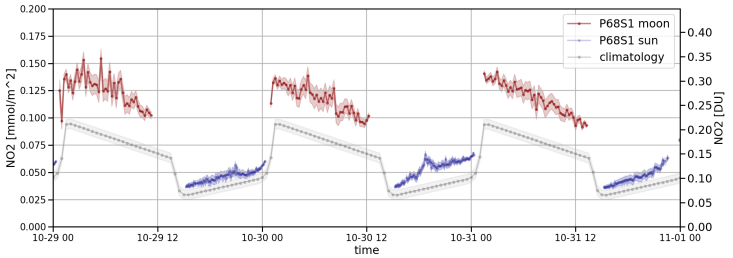


Figure 72: JPL measurements plotted with the stratospheric climatology in 72a with full moon on 24th October 2018 and Pandora measurements with the climatology with full moon on 31st October 2020.



(a) JPL measurements as mentioned in [34]. The data is publicly available.



(b) Direct sun (5min averages) and direct moon (15min averages) NO_2 for WrightwoodCA with the stratospheric climatology (half hourly steps).

4.3 Direct sun O_3 temperature validation

4.3.1 External validation

Comparison datasets

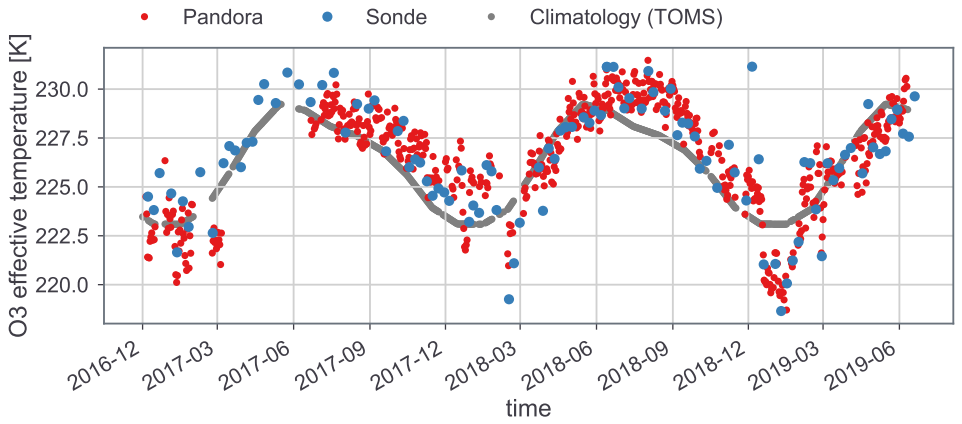
Table 9: Comparison dataset(s) for total column O_3temp validation.

Instrument	Date	PI
Izana, Pandora 121s1		
ECC Ozonesonde	2016-12-1 to 2019-6-10	Carlos Torres, Alberto Redondas

Izana

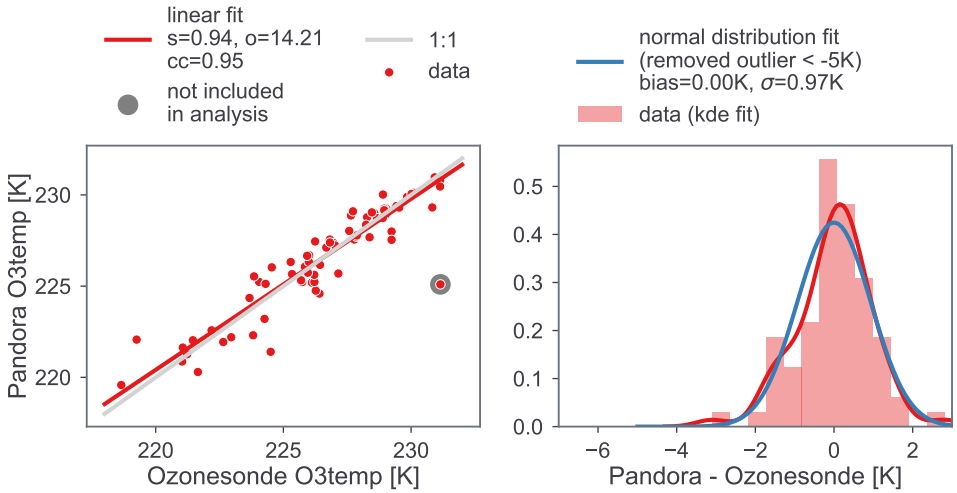
The times series of the O_3temp retrieved from Ozonesondes is shown in Figure 73 (as blue dots) and compared to diurnal mean values of O_3temp retrieved from the Pandora (red dots). Regarding the Ozonesonde retrieval it needs to be said that a crucial step is the extrapolation of the sonde profile to about 60 km height. Burst heights for sondes hardly exceed 35 km. At that altitude still significant O_3 partial pressure can be measured, which makes the need to extrapolate the sonde profile data to close the gap to column measurements. The dataset used here is extrapolated with climatological O_3 profiles collected at Izana, scaled to the respective sonde profiles (work thankfully done by Alberto Redondas from AEMET).

Figure 73: O_3temp times series in Izana, derived from ECC Ozonesonde (blue) and Pandora (red). Pandora data are diurnal means. For comparison, O_3temp derived from the TOMS v8 ozone profile climatology is also shown (in gray).



The analysis of the difference of both data sources (Pandora O_3temp are diurnal means) unfolds a rather high agreement, as illustrated in Figure 74, with a vanishing mean bias and a 1σ standard deviation of about 1 K (right figure panel). Further, the linear regression (left figure panel) reveals a slope close to unity (Pandora retrieved O_3temp appears to slightly overestimate) and a very high correlation of 0.95. Please note that one Ozonesonde data point was not included in the analysis (gray shaded dot in the left figure panel), as the values seems to be an outlier (cross-check with time series in Figure 73, Dec. 5th, 2018.)

Figure 74: Comparison of O_3temp as retrieved from ECC Ozonesonde and Pandora, based on Izana data from 2017 to 2019.

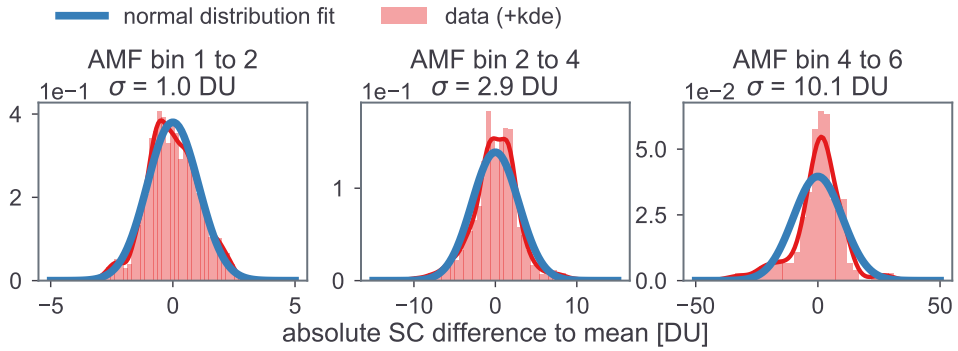


4.4 Direct sun total O_3 validation

4.4.1 Internal validation

For InterComp we utilize the already introduced Greenbelt triad (Pandora 66s1, 67s1 and 68s1). If the distribution of the absolute SC difference to the mean is calculated for the combined Pandora dataset, a 1σ deviation calculates to 1 DU for AMFs between 1 and 2 (Figure 75, left panel), 3 DU for AMFs between 2 and 4 (middle panel) and reaches 10 DU for even higher AMFs. This gradient is very likely linked to SpecSL differences and the offset in the retrieved O_3temps.

Figure 75: Distribution of the absolute SC differences to the mean for the combined Pandora datasets. In the thee figure panels, the distribution is shown for AMF bin 1 to 2 (left), 2 to 4 (middle) and 4 to 6. The actual data are depicted in red (with a KDE fit in dark red) and a normal distribution fit in blue.



4.4.2 External validation

Comparison datasets

Table 10: Comparison dataset(s) for total column O_3 validation.

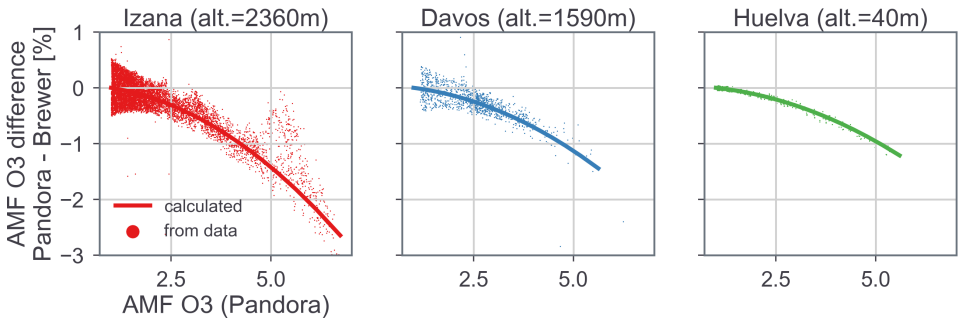
Instrument	Date	PI
Izana, Pandora 121s1		
Brewer #157	2016-12-1 to 2019-6-10	Alberto Redondas
Brewer #183	2016-12-1 to 2019-6-10	Alberto Redondas
Brewer #185	2016-12-1 to 2019-6-10	Alberto Redondas
Davos, Pandora 120s1		
Brewer #163	2020-8-6 to 2020-12-16	Julian Gröbner
Huelva, Pandora 121s1		
Brewer #150	2019-6-18 to 2019-6-28	Jose Manuel Vilaplana
Brewer #183	2019-6-18 to 2019-6-28	Alberto Redondas
GreenbeltMD, Pandora 68s1		
Brewer #171	2019-3-10 to 2019-10-1	Gordon Labow

Izana, Davos, Huelva, GreenbeltMD

In principle we would like to compare vertical total column amounts of O_3. However, the calculation of the direct sun AMF is done slightly differently for Brewers and Pandoras, leading to an AMF dependent difference (Pandora AMF higher). This is exemplarily shown for three example time series (Izana, Davos, Huelva), covering different latitudes and station altitudes in Figure 76, where the percentage AMF difference is shown as calculated (for one scenario, solid lines) or directly from the data (dots). The two main differences in AMF calculations are:

- The effective O_3 layer height is set constant to 22 km for the Brewer AMF, but is adjusted according to the O_3 profile climatology for Pandora AMF as a function of season, latitude and column amount.
- In contrast to the Brewer AMF, the Pandora AMF calculation increases the effective O_3 layer height further by the station altitude.

Figure 76: O_3 AMF difference between Pandora and Brewer retrievals (as explained in the text). The three figure panels show examples from Izana (left), Davos (center) and Huelva (right). Differences calculated from measurements are shown as dots and the lines represent the AMF model calculations.



As a consequence we have performed the dataset comparison based on slant columns and not on vertical columns. Figure 77 summarizes the comparison results, showing the percentage difference Pandora - Brewer as a function of AMF for the two operational O_3 products: The product using a synthetic reference and does O_3temp fitting (left figure panels) and the product using the literature references with O_3temp climatology (right figure panels). Each figure row refers to a different comparison location. In case several Brewers have been available at a location (compare Table 10), the different results are shown in different colors in each figure panel. Further, the correlation coefficient (abbreviated cc) and the slope of a linear regression (abbreviated s) are listed in the figure legends. To enhance the perceptibility, we have overlain AMF binned mean values (large dots), connected with solid lines.

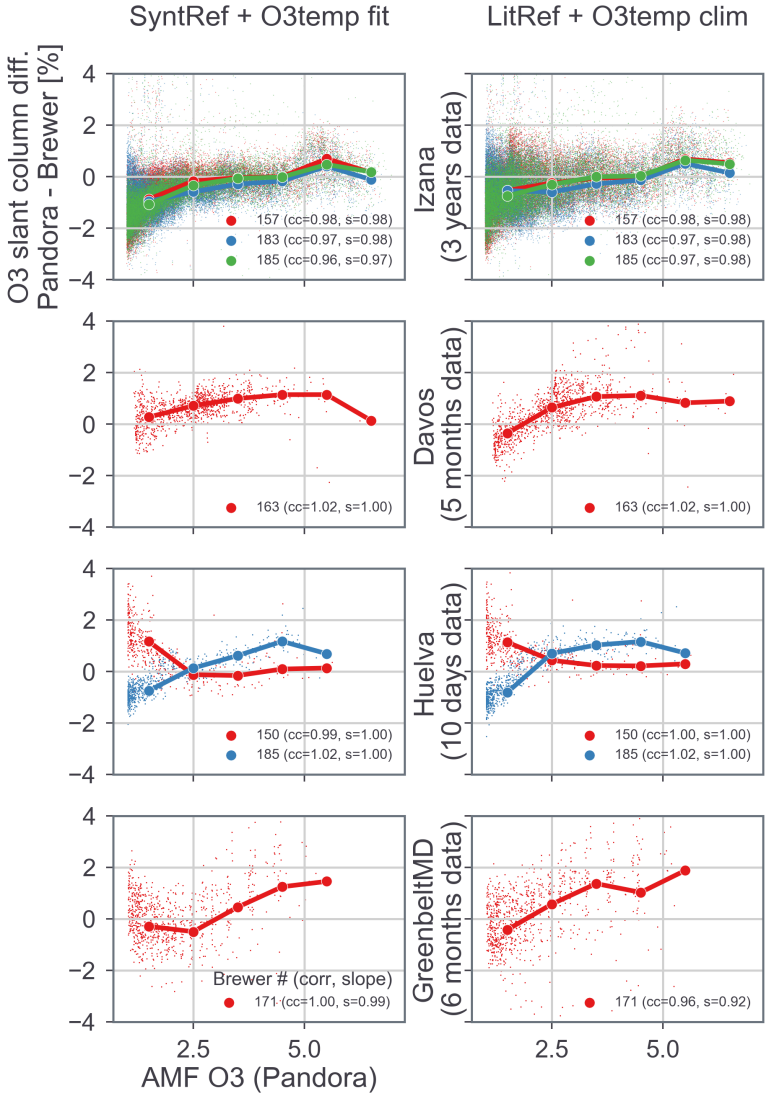
The analysis of the difference reveals a very high agreement between Pandora and Brewer with correlation coefficients and regression slopes close to unity for all cases, even for the three years dataset of Izana.

Both operational data products perform similarly in terms of offset and AMF dependency, with an slightly enhanced data scatter for the product using the literature reference (as expected).

The Pandora data tendentially show higher values at higher AMF as the Brewer, leading to a slight AMF dependent difference in most cases. This can be partly explained by the non-linear cross-section model used in the Blick processor, which, in compensation of non-linear O_3 absorption, gradually increases slant column values by up to about 1 % at AMF 5.

At smallest AMFs, potential differences due to O_3 calibration are most pronounced in this percentage view. Here we see e.g. an average calibration difference of -1% at Izana (the EuBrewNet reference Brewers) good agreement at Davos and GreenbeltMD and different biases against the two Brewers Brewers in Huelva. Overall the agreement to the Brewer resides within $\pm 2\%$ across the hole AMF range.

Figure 77: Relative O_3 slant column differences between Pandora and Brewer estimated for different locations (figure rows) and Brewers (individual colors in figure panels). The left column refers to the O_3 product based on a synthetic reference (SyntRef) and the right column to the O_3 product based on a literature reference (LitRef). Correlation coefficient (cc) and the slope of a linear regression (s) are given in the respective legends. AMF binned mean values for each case are shown as large dots, connected with solid lines.



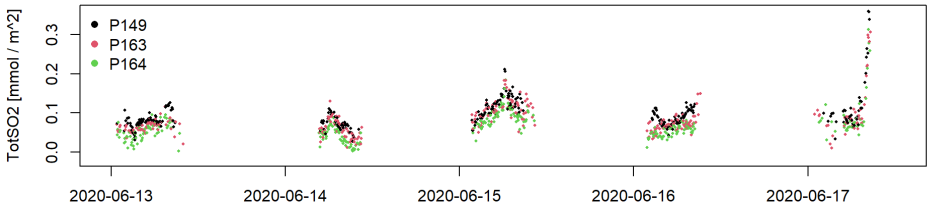
4.5 Direct sun total SO_2 validation

4.5.1 Internal validation

The InterComp utilizes the direct sun datasets of three Pandoras (P149, P163 and P164), co-located at Seoul, Republic of Korea, from 2020-06-06 to 2020-06-21. All three units have no matrix straylight correction included, which limits the AMF validity to 3. The analysis of the three datasets follows the statistical framework outlined in Appendix D.

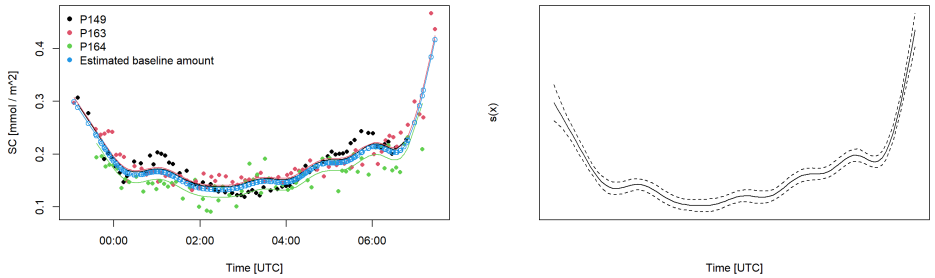
Example days of total SO_2 are illustrated in Figure 78. The instruments follow common patterns, although there is also the indication that P164 is systematically lower than the other two instruments.

Figure 78: Timeseries of total column SO_2 for Pandoras P149, P163, P164 at Seoul-SNU from 2020-06-13 to 2020-06-17



The GAM approach is applied on the SO_2 slant columns for each day, as exemplarily shown in Figure 79. The instruments share a daily effect, which can be captured by the regression model, as seen in the right figure panel. However, since P149 shows two peaks symmetrically around noon, which are not well-expressed for the other Pandoras, the daily effect at these hours is giving more weight to the P163 and P164.

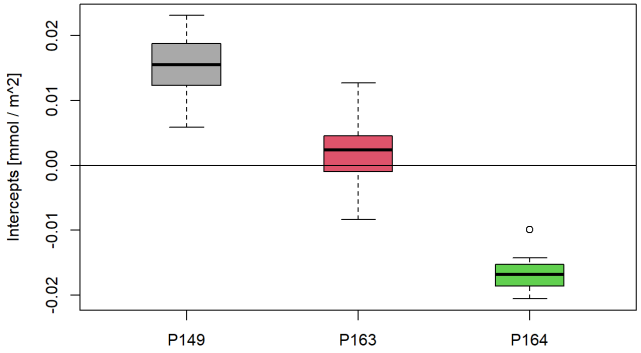
Figure 79: Example day of the daily effect s(x) (right graphic). Left graphic shows the measured slant columns as filled dots. Additionally, the estimated baseline amount is illustrated as blue circles.



An evaluation of the intercepts over all days does show systematic calibration errors on median for the individual instruments (Figure 80), where P164 is as expected 0.0168 $mmol/m^2$ lower than the baseline, and P149 0.0155 $mmol/m^2$ larger. Therefore, the range of median intercepts is 0.0324 $mmol/m^2$, which can be seen as an empirical calibration error estimate since one cannot know which instrument is closer to the truth.

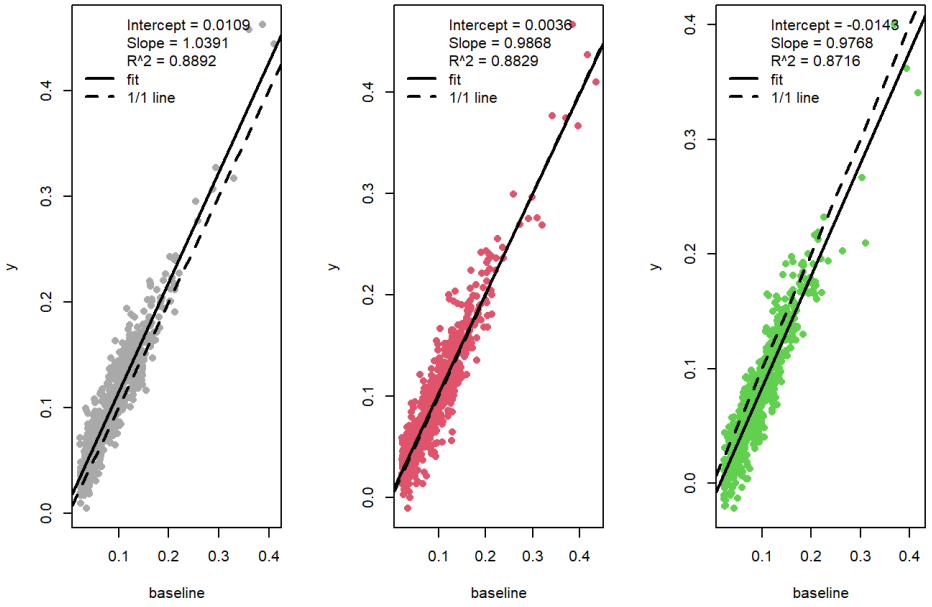
This enhanced uncertainty reflects the air mass factor limit of 2.5-3 which had to be applied for those Pandora units.

Figure 80: Intercepts of the obtained daily GAM fits, relative to the baseline intercept



The scatter plot in Figure 81 reports similar intercepts, and shows R^2 values between 0.87-0.88.

Figure 81: Scatterplots for the three instruments showing the baseline SO_2 slant column in $[mmol/m^2]$ (x-axis) against the measured value (y-axis). The legend gives the regression coefficient for a linear fit, and the R^2 value. Color-coding as in Fig.80, from left to right: P149, P163, P164.

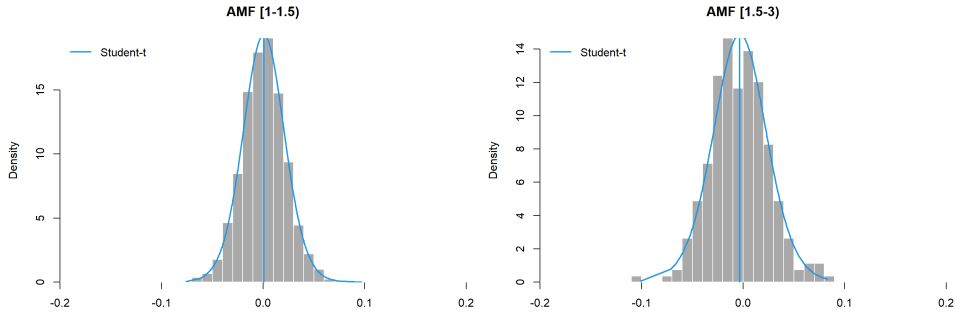


The corresponding error distribution for all instruments together compared to the estimated baseline amount, is illustrated in Figure 82, and the corresponding statistics are presented in Table 11. The inner 68% interval reports an uncertainty ranging from +/- 0.020 to +/- 0.027 $mmol/m^2$.

Table 11: Summary statistics of the Student's t distribution for the AMF errors shown in Figure 82. From left to right: Location parameter (Location), scale parameter (Scale), degree of freedom (df), inner 68% interval width (PI68), inner 95% interval width (PI95).

	Location	Scale	df	PI68	PI95
AMF [1-1.5)	0.00053	0.02042	21.96723	0.04155	0.08469
AMF [1.5-3)	-0.00337	0.02638	16.02136	0.05414	0.11182

Figure 82: Error distribution comparing the measured slant columns of all instruments with the baseline amount, evaluated for different AMF categories. X-axis denotes the error in [mmol / m^2].



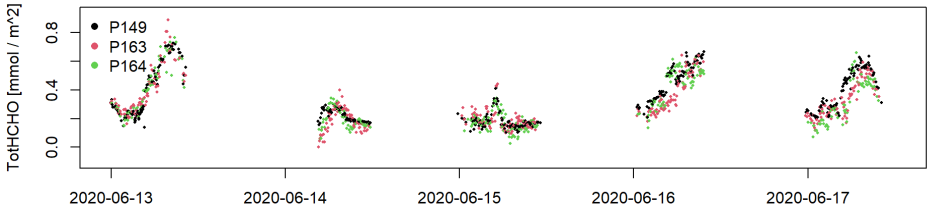
So far no datasets to conduct an external validation are available.

4.6 Direct sun total HCHO validation

4.6.1 Internal validation

The InterComp is done based on the same dataset and analysis as explained in 4.5.1. Example days of total HCHO are illustrated in Figure 83.

Figure 83: Timeseries of total column HCHO for Pandoras P149, P163, P164 at Seoul-SNU from 2020-06-13 to 2020-06-17



The GAM approach is applied on the HCHO slant columns for each day, as illustrated in Figure 84. The instruments share a variable daily effect, which can be captured by the regression model, as illustrated in Figure 84 (right). An evaluation of the intercepts over all days does not reveal a significant systematic calibration error on median for the individual instruments (Figure 85), although P164 indicates to be slightly lower than the other two instruments. However, the range of median intercepts is 0.0079 $mmol/m^2$, which is an indicator that the calibration with the E-MLE approach is applicable.

Figure 84: Example day of the daily effect s(x) (right graphic). Left graphic shows the measured slant columns as filled dots. Additionally, the estimated baseline amount is illustrated as blue circles.

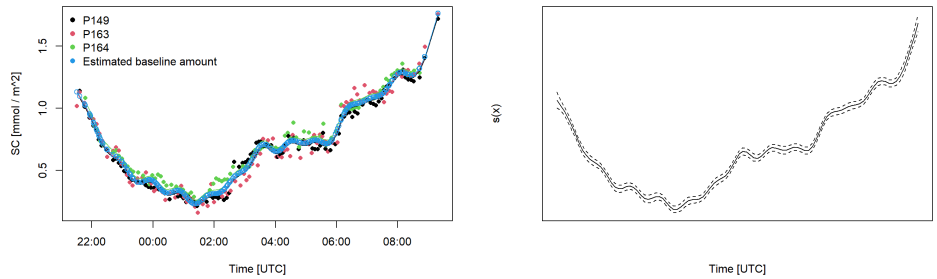
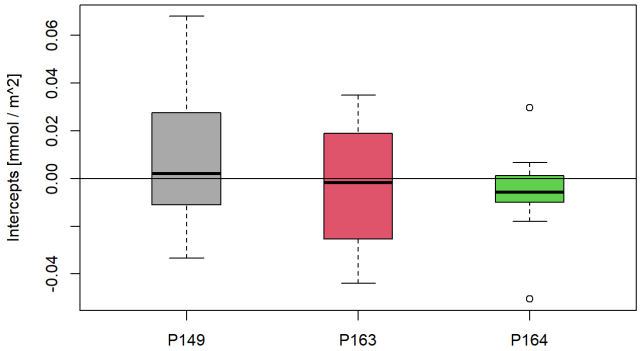
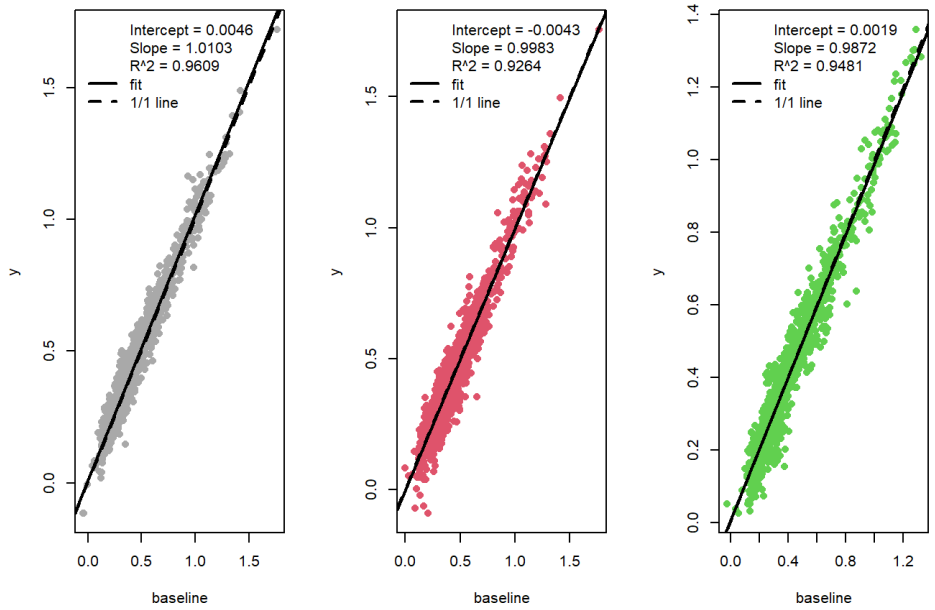


Figure 85: Intercepts of the obtained daily GAM fits, relative to the baseline intercept



The scatter plot in Figure 86 is supporting the intercept evaluation with its linear intercept close to zero, and additionally shows R^2 values between 0.92-0.96.

Figure 86: Scatterplots for the three instruments showing the baseline HCHO slant column in $[mmol/m^2]$ (x-axis) against the measured value (y-axis). The legend gives the regression coefficient for a linear fit, and the R^2 value. Color-coding as in Fig.85, from left to right: P149, P163, P164.



The corresponding error distribution for all instruments together compared to the estimated baseline amount, is illustrated in Figure 87, and the corresponding statistics are presented in Table 12. The inner 68% interval reports an uncertainty ranging from +/- 0.048 to +/- 0.057 $mmol/m^2$. Since the calibration divergence is found to be 0.008 $mmol/m^2$ (see above), the major driver of uncertainty is expected to come from temporal spectral disturbances (smooth and structured), associated with bad FOV and pointing. This is accounted to the so-called structured uncertainty.

Table 12: Summary statistics of the Student's t distribution for the AMF errors shown in Figure 87. From left to right: Location parameter (Location), scale parameter (Scale), degree of freedom (df), inner 68% interval width (PI68), inner 95% interval width (PI95).

	Location	Scale	df	PI68	PI95
AMF [1-2)	0.00059	0.05542	13.65305	0.11439	0.23830
AMF [2-3)	0.00016	0.04500	7.03352	0.09628	0.21259
AMF [3-6)	0.00958	0.05313	12.36208	0.11009	0.23075

The CrossComp is performed for Pandora units covering at least half a year of both direct sun and multi axis observations, and being Delrin-free. The statistical framework as described in Appendix D.2 is adjusted by using only 2 intercepts, but still a shared daily effect. Due to the different viewing geometry leading to total and tropospheric column HCHO data products, we expect less of an agreement, in particular if the observed species is not homogeneously distributed in the atmosphere. Therefore, the focus of the evaluation is on the systematic difference, which is quantified by the intercepts.

Figures 89, 90, 91 show two example days, each for MexicoCity (P142), Wakkerstroom (P159), and Seoul (P164). In the shown examples, all three locations illustrate a daily effect which is similar for both the total vertical HCHO columns from direct sun, and the tropospheric vertical column amounts from sky data. Until the HCHO outgasing of the sensor head has been solved, such an agreement was not visible (Figure 32).

Figure 89: HCHO for P142 in MexicoCity, from direct sun (black dots) and from multi axis observation (red). Lines represent the fitted daily effect as defined in D.2, with the blue dots showing the estimated baseline amount.

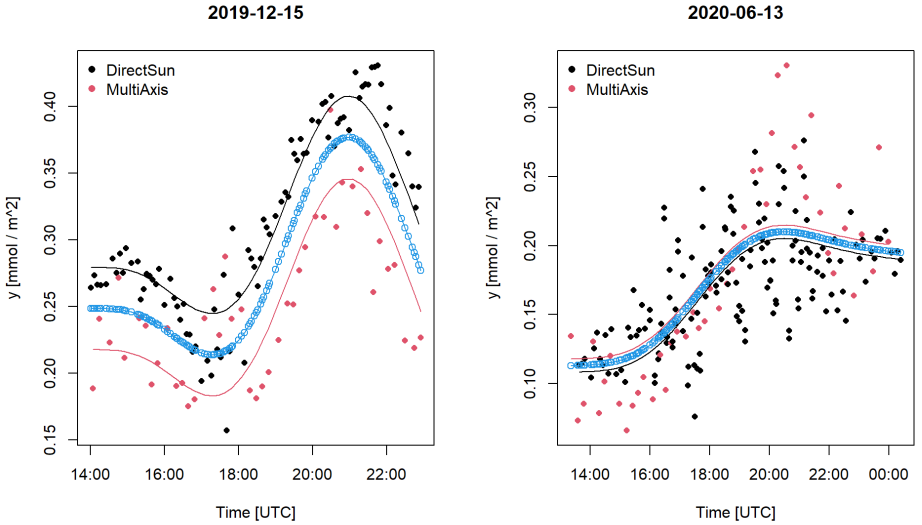


Figure 87: Error distribution comparing the measured HCHO slant columns of all instruments with the baseline amount, evaluated for different AMF categories. X-axis denotes the error in [mmol / m^2].

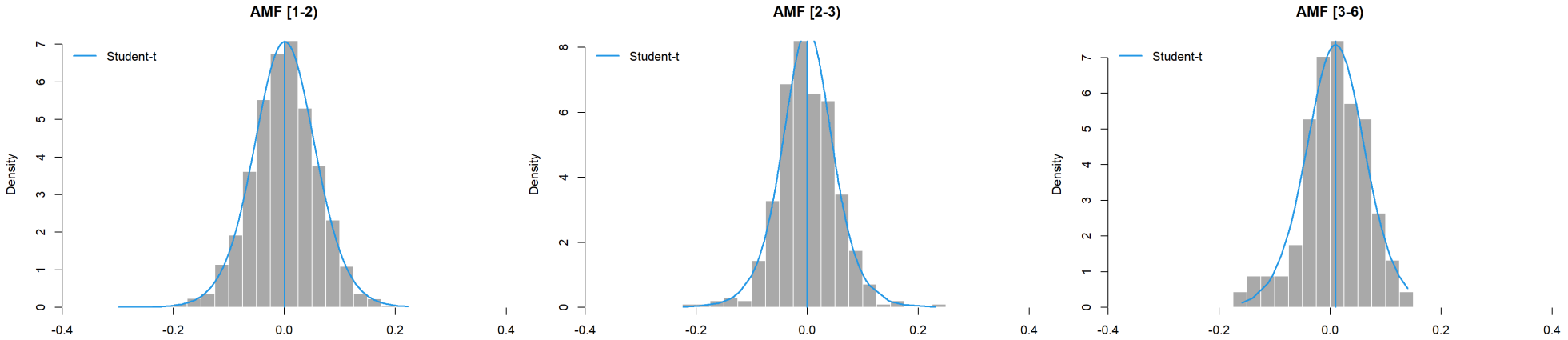


Figure 88: GAM intercepts of multi axis measurements relative to direct sun measurements, for P142 (MexicoCity), P159 (Wakkerstroom, South Africa), P164 (Seoul, Republic of Korea). Boxplot (left graphic) represents intercepts for all days, which are illustrated in the right figure as a function of the day of the year.

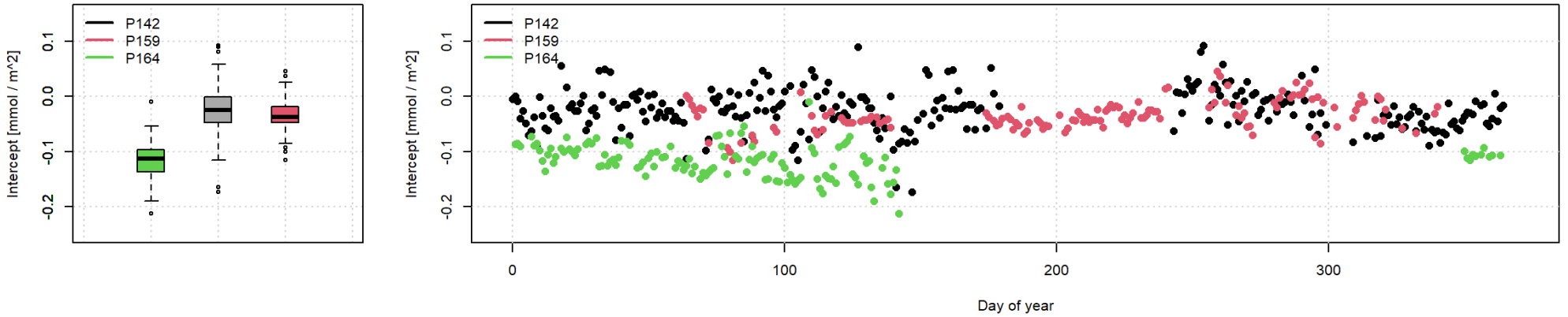


Figure 90: As in Figure 89, but for P159 in Wakkerstroom, South Africa.

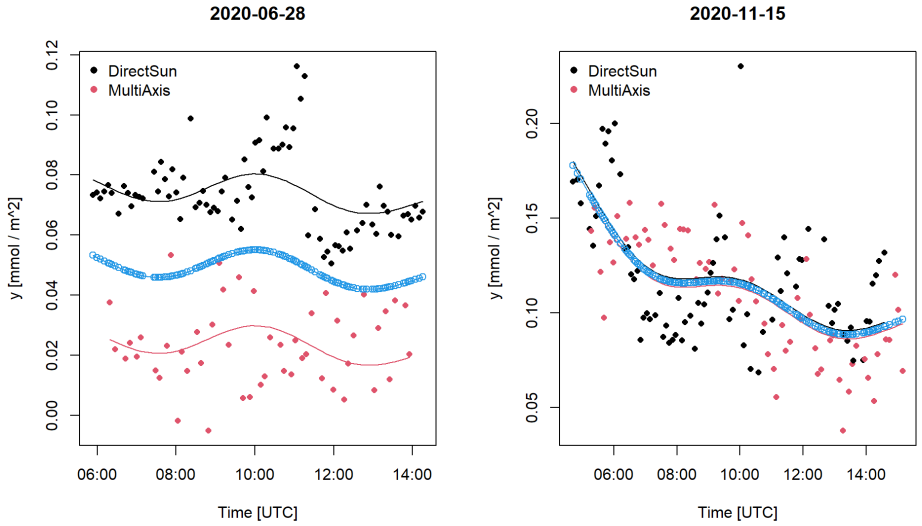
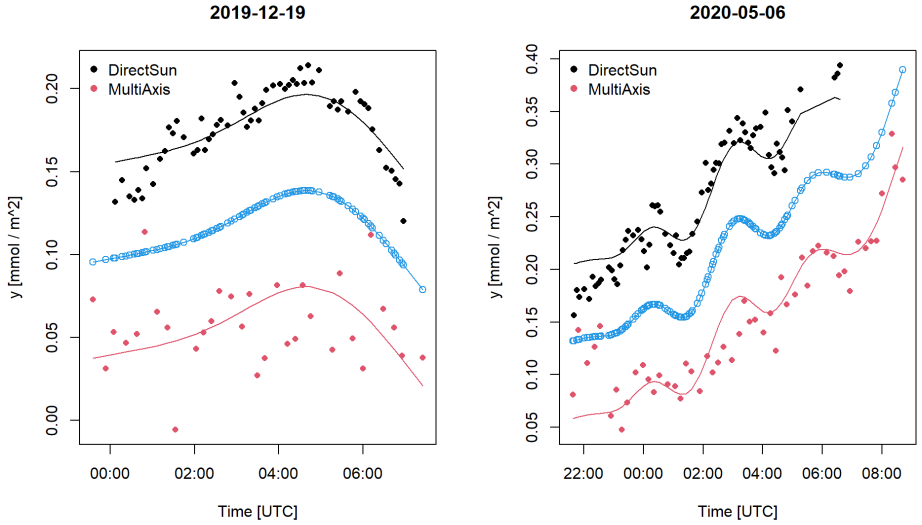


Figure 91: As in Figure 89, but for P164 in Seoul, Republic of Korea.



However, these examples highlight an offset, with the tropospheric column amounts being systematically lower than the total vertical column amounts from direct sun observations. The negative intercepts relative to direct sun measurements obtained by the GAM are illustrated in Figure 88. P142 implies a seasonal dependence of the intercepts, but without multiple years and other instruments covering more than two seasons, this cannot be generally concluded.

It has to be mentioned, that this systematic difference does not necessarily refer to a calibration error. First, tropospheric columns are based on sequential reference spectra[8]. Second, if the reference value of the total columns would have been guessed incorrectly, an AMF dependent bias would be introduced. Therefore, this difference is most likely be associated with residual HCHO layer(s) above the detection height of the sky data product. According to comprehensive analysis of sky data based products, *Tirpitz et al.* [51] concluded that species above about 2 km height cannot be reliably retrieved due to the strong decrease of sensitivity with height of sky data. Consequently, it can be expected that absorbers above the lower troposphere can only be seen in direct sun measurement mode. A similar pattern was observed even at the high altitude site of Jungfraujoch, Switzerland comparing FTIR and MAX-DOAS [24]. This might raise the question whether sky data products from ground allow a comprehensive validation of satellite derived HCHO.

4.6.2 External validation

Comparison datasets

Table 13: Comparison dataset(s) for total column HCHO validation.

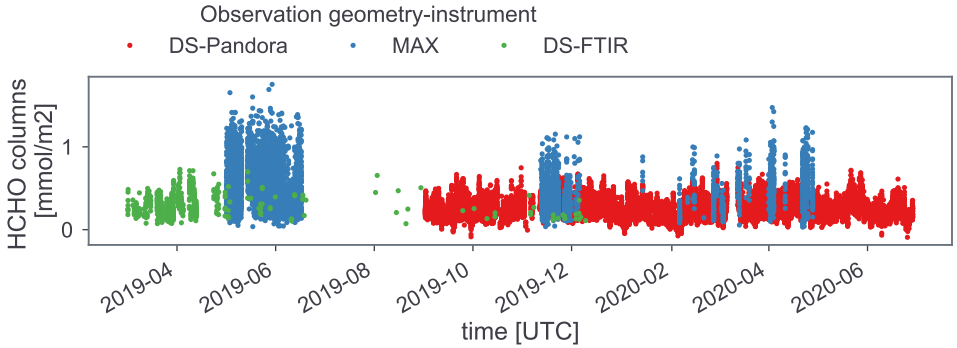
Instrument	Date	PI [Publication]
	Mexico City, Pandora 142s1	
MAX-DOAS	2019-5-1 to 2020-4-15 (not continuous)	Claudia Rivera Cárdenas [18].
FTIR	2019-3-1 to 2019-12-1 (not continuous)	Claudia Rivera Cárdenas [18].

[8]Each measurement sequence uses its own reference spectrum. In our case a sequence takes about 3 minutes. Across this time span the instrument can be expected to be perfectly stable which makes all instrumental features to cancel out.

Mexico City

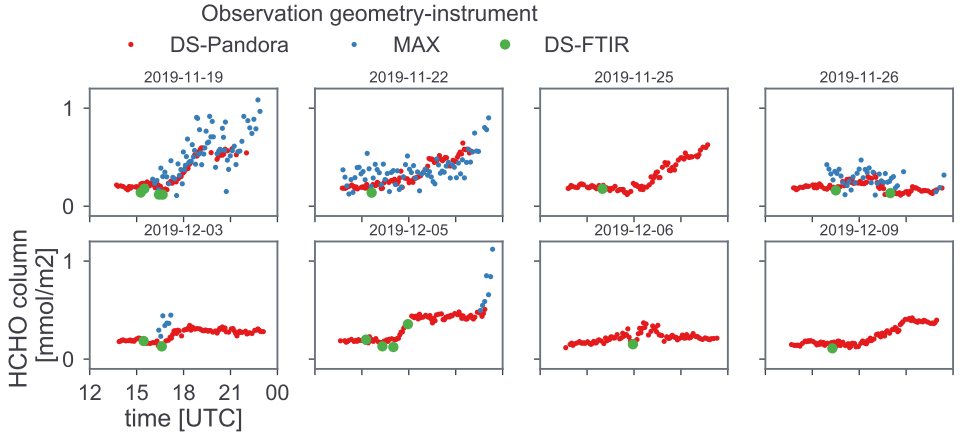
The time series of the available comparison datasets in Mexico City (see Table 13) is shown in Figure 92 (MAX-DOAS in blue, FTIR in green), together with the Pandora data (in red).

Figure 92: Timeseries of columnar HCHO for Mexico City from March 2019 to July 2020. Shown are total columns from Pandora (red) and FTIR (green) and tropospheric columns from MAX-DOAS (blue).



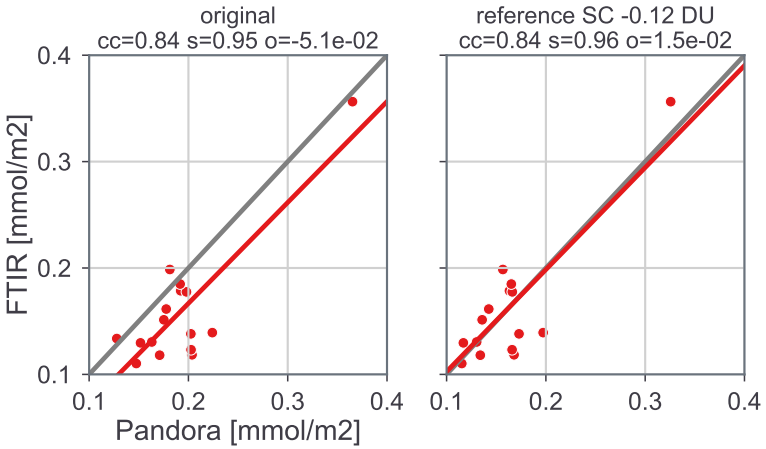
A close up in the period around November/December 2019, shown in Figure 93, reveals a rather large scatter of the MAX-DOAS dataset compared to Pandora vs FTIR. A corresponding correlation coefficient of only 0.3 makes us to question the data quality of the MAX-DOAS dataset and therefore we decided to limit the further analysis to comparisons to the FTIR instrument only.

Figure 93: Same figure description as for Figure 92, however for a close- up of the time series.



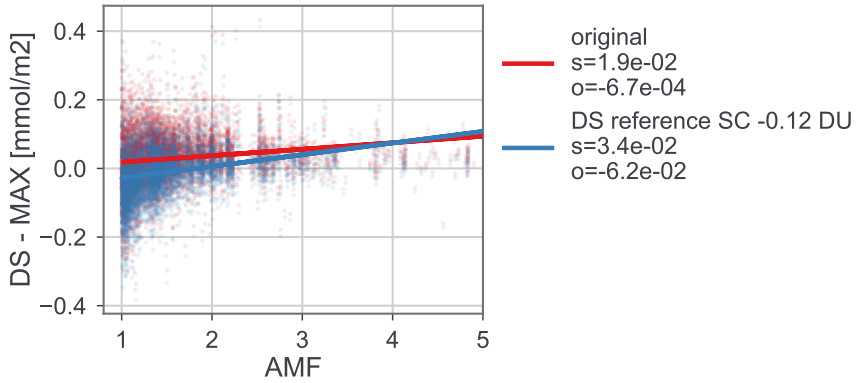
Although the short overlap period for the FTIR and the Pandora only allows to consider 16 data points, we still carried out a basic analysis. Figure 94 displays the correlation between Pandora (x-axis) and the FTIR (y-axis) in a scatter plot, with the comparison results for the original case shown in the left figure panel. A correlation 0.84 and a slope close to unity (0.95) indicate a rather high relative agreement. However, there appears to be an additive SC bias of -0.12 DU (-0.05 mmol m^{-2}), visible in the right figure panel.

Figure 94: Comparison of Pandora and FTIR DS total column HCHO. Left figure panel shows the original data. For the comparison in the right figure panel, a Pandora SC additive bias of -0.12 DU is assumed.



As a test, we applied this apparent bias to the Pandora DS data, that is subtracting 0.12 DU from the Pandora DS SC. However, if this is done, we worsen the internal comparison to the tropospheric columns from the MAX mode, as illustrated in Figure 95: For the original dataset (red dots with linear fit as red line), there is only a rather small AMF dependency discernible for the difference of the HCHO columns from DS and MAX. This AMF dependency is increased (blue dots with linear fit as blue line) if the bias correction is applied.

Figure 95: AMF dependency of the difference HCHO columns from DS and MAX data from Pandora. The red dots represent the original case (see Figure 94) with a slope of the linear fit being smaller as for the case with the bias corrected DS data (blue colors).



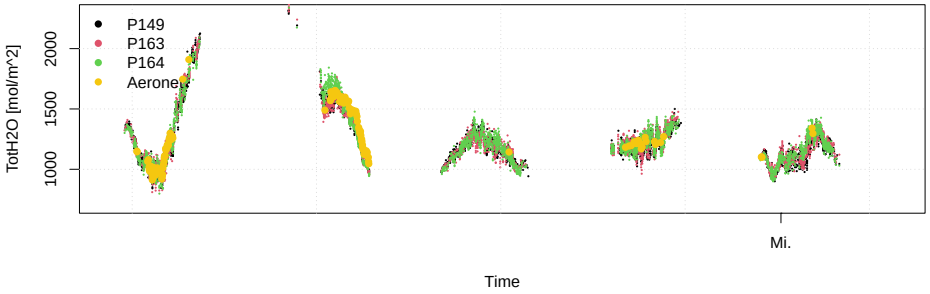
Assuming that the MAX column retrieval does not suffer from a significant AMF dependent intrinsic error, we cannot necessarily address the bias to the Pandora. In particular, since we have to be aware about the fact that FTIR HCHO retrievals firstly have to assume an a priori vertical HCHO profile for the minimization and secondly use micro windows in the IR wavelengths (≈3.6 μm), usually based on HITRAN libraries.

4.7 Direct sun total H_2O validation

4.7.1 Internal Validation

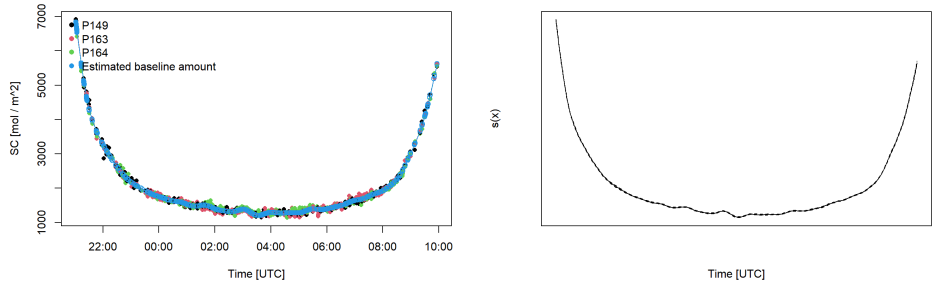
The InterComp is done based on the same dataset and analysis as explained in 4.5.1. Figure 96 gives an overview of the retrieved total column H_2O amounts for Pandora 149, 163, 164, all individually calibrated. Additionally, an Aeronet site is available and added for the overview, which is also compared to in the external validation section.

Figure 96: Timeseries of total column H_2O for Pandoras P149, P163, P164 at Seoul-SNU from 2020-06-12 to 2020-06-18. Filled yellow circles illustrate data from the Aeronet site.



The GAM approach is applied on the H_2O slant columns for each day, as illustrated in Figure 97. The instruments share a variable daily effect, which can be captured by the regression model, as illustrated in Figure 97 (right). An evaluation of the intercepts over all days does not reveal a significant systematic calibration error on median for the individual instruments (Figure 98). The range of median intercepts is 6.5 mol/m^2, which is an strong indicator that the calibration with the MLE is suitable.

Figure 97: Example day of the daily effect s(x) (right graphic). Left graphic shows the measured slant columns as filled dots. Additionally, the estimated baseline amount is illustrated as blue circles.



This agreement is also supported in the good correlations statistics as illustrated in Figure 99, where both R^2 values and slope coefficients are close to 1.

Figure 98: Intercepts of the obtained daily GAM fits, relative to the baseline intercept

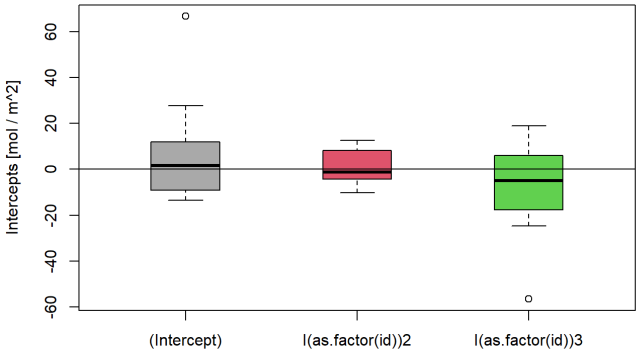
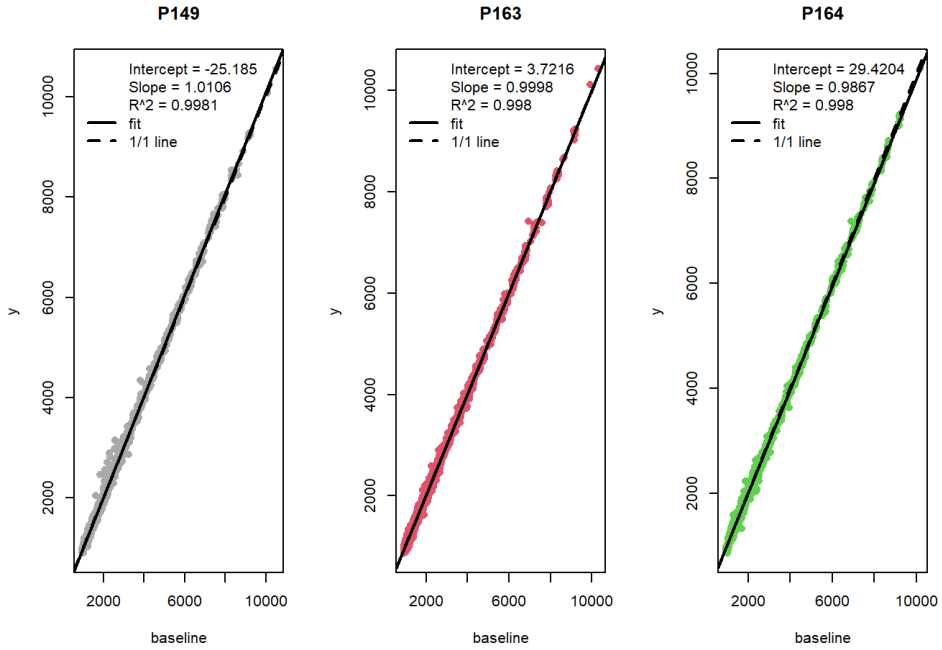


Figure 99: Scatterplots for the three instruments showing the baseline H_2O slant column in $[mol/m^2]$ (x-axis) against the measured value (y-axis). The legend gives the regression coefficient for a linear fit, and the R^2 value. Color-coding as in Fig.98, from left to right: P149, P163, P164.



The error distribution for all instruments together is illustrated in Figure 100, with the corresponding summary statistics shown in Table 14. The inner 68% interval ranges from +/- 48.828 to +/- 66.785 mol/m^2.

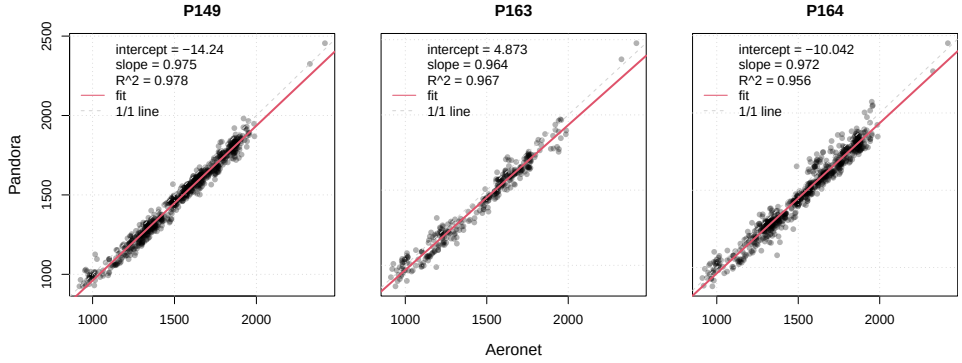
Table 14: Summary statistics of the Student's t distribution for the AMF errors shown in Figure 100. From left to right: Location parameter (Location), scale parameter (Scale), degree of freedom (df), inner 68% interval width (PI68), inner 95% interval width (PI95).

	Location	Scale	df	PI68	PI95
AMF [1-2)	0.03292	45.58462	6.92719	97.65515	216.04118
AMF [2-3)	-0.88677	47.98399	7.57297	102.12680	223.49366
AMF [3-7)	-1.01987	63.64907	9.50481	133.57030	285.65382

4.7.2 External Validation

The PGN location Seoul-SNU has an Aeronet site which can be used for the overlap period of the three Pandoras 149, 163, 164. An example period is shown in Figure 96. The time interpolated Pandora measurements show high R^2 values for all instruments above 0.95 with a small negative bias for P149 and P164, as visible in the figure legends of Fig. 101.

Figure 101: Scatterplot for Aeronet H_2O (x-axis) against interpolated Pandora H_2O (y-axis) for P149, P163, P164 from left to right.



Although the correlation is very high already, there is very likely some comparison "noise" added, considering the different time stamps (and corresponding interpolation) and the highly variable H_2O concentrations in time.

Two more Aeronet sites (Juelich and Davos) are compared in Figure 102, which both have a long overlapping timeseries (Figure 51) of roughly 2 years. Both sites report a very high R^2 value of 0.99 with an average difference less than $-10 mol/m^2$.

Also the use of the Kurucz reference spectra,which leads to the so-called 'out of the box' H_2O with out the need of a field calibration, shows a promising aggreement with the corresponding Aeronet sites at Davos an Juelich (Figure 103) with R^2 values of 0.99.

Figure 100: Error distribution comparing the measured H_2O slant columns of all instruments with the baseline amount, evaluated for different AMF categories. X-axis denotes the error in [mol/m^2].

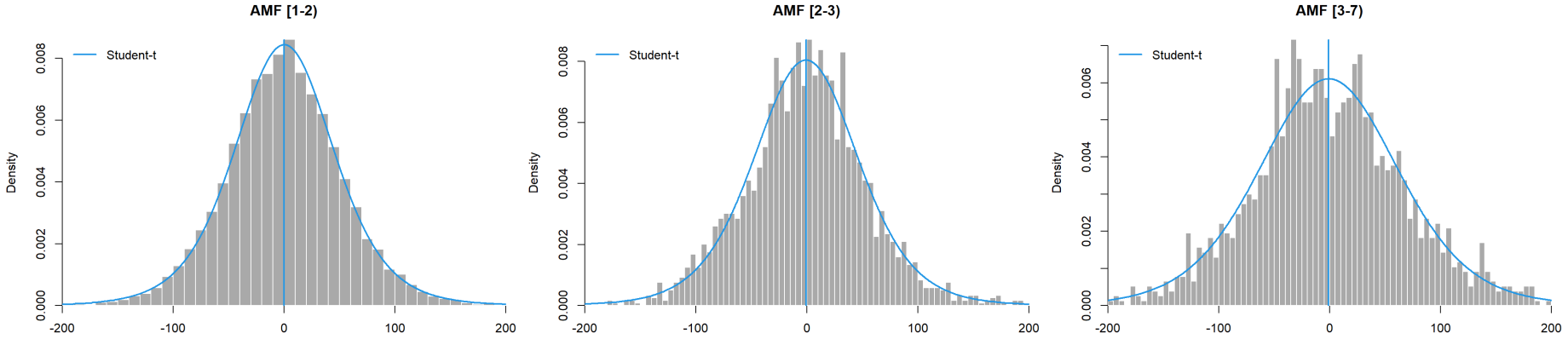


Figure 102: Scatterplot for Aeronet H_2O (x-axis) against interpolated Pandora H_2O (y-axis) and difference for P30 (top row) and P120 (bottom row) using the synthetic reference.

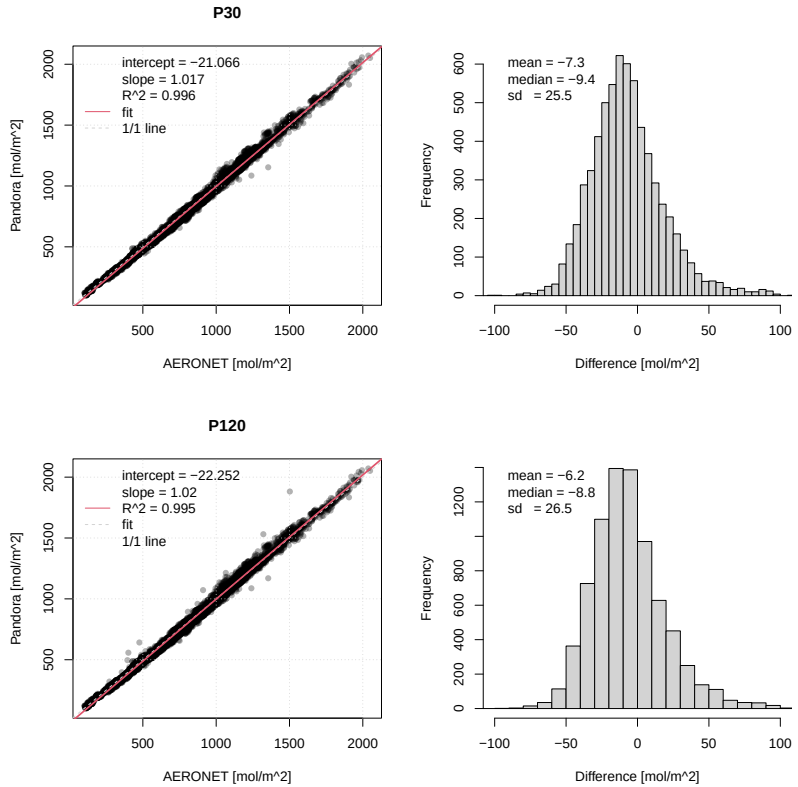
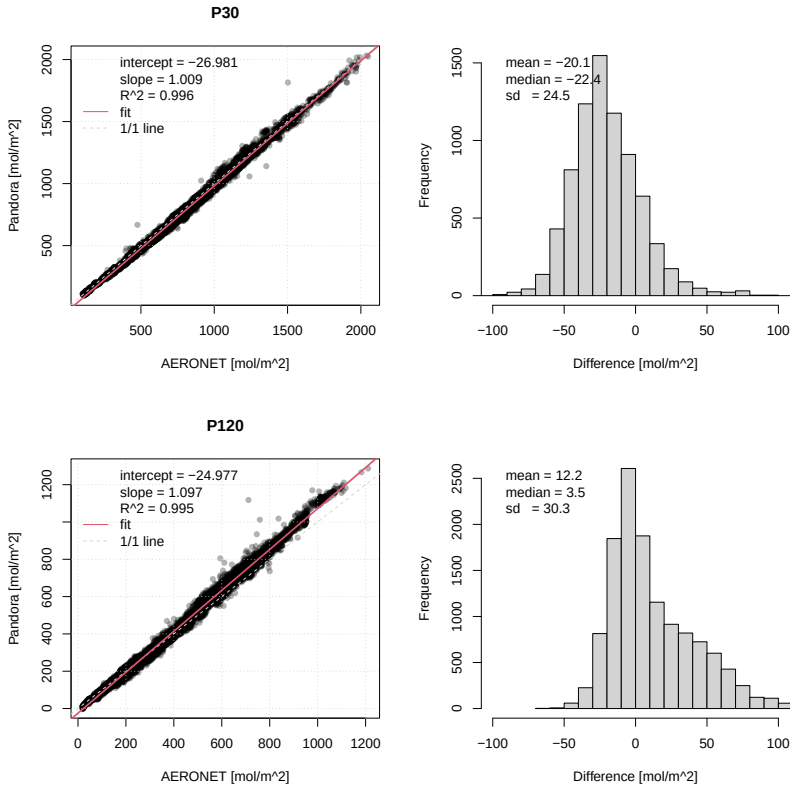


Figure 103: Scatterplot for Aeronet H_2O (x-axis) against interpolated Pandora H_2O (y-axis) and difference for P30 (top row) and P120 (bottom row) using the extraterrestrial reference.



The apparent ≈ 4% underestimation for the Seoul Pandoras compared to Aeronet should be studied further as it is not straightforward to address: Aeronet retrieves H_2O columns from the very non-linear and strongly temperature dependent 940 nm band. In contrast, the Pandora data are based on a rather linear band with hardly any temperature dependency. Both components, however, are known to have an impact on the SC amount and AMF dependency. Although these points are argued to be considered well in the Aeronet algorithm [38], further investigations would be advisable.

5 Current main reasons for development limitations

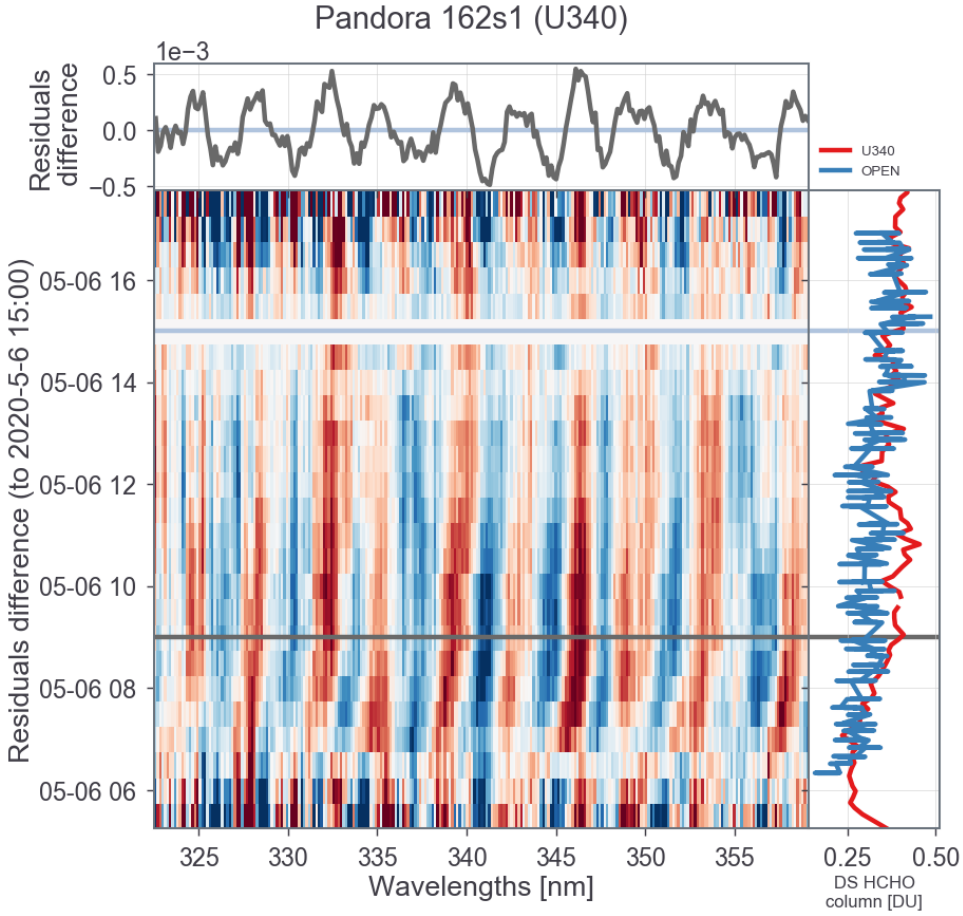
The development of some data products had to be temporarily discontinued due to recently identified and partially occurring instrumental limitations of the following types:

Filter etalon (LimUVFilt) Retrievals in the UV regions are based on measurements utilizing a band-pass filter sandwiched with a diffuser. So it is actually two optical filters mounted next to each other. Under yet unidentified circumstances, this filter combination appears to produce spectral etalon structures that interfere and bias the spectral fitting algorithms. Whilst this impact is mostly tolerable for the retrieval of weak absorber like HCHO, the retrieval of very weak absorber like HONO is already significantly distorted.

One example, where the feature is even impacting HCHO, is given in Figure 104. This figure shows one day for Pandora 162s1 in Brussels and consists of three parts. On the very right the flipped time series (y axis is time) for the column amount is shown when a band-pass filter sandwich is used (U340=red) or no filter is used (OPEN=blue). The spectral residuals relative to a time where measurements of both filter types agree (faint blue horizontal line), is visualized as heat map. A clear correlation can be seen between biased column amounts and the magnitude of the residual difference. The etalon like structure gets obvious (top panel in the figure) when looking at the relative residuals where the deviation of the column data is large (gray horizontal line).

This etalon is expected to vanish if only one optical element would be used. In order to test this, we conceived a study to investigate measurements based on a custom home-made (SciGlob) band-pass filter what has been polished on one side (to act as diffuser). This study is still ongoing.

Figure 104: Total column HCHO variation for retrievals using a band-pass filter (red) or without a filter (blue) is shown on the very right for Pandora 162s1 in Brussels. The change in the spectral residual, relative to the time stamp marked by the blue line (least difference in column amount) is reflected in the heat map. This residual difference shows a clear etalon pattern when the column difference is large (gray horizontal line and top figure panel).

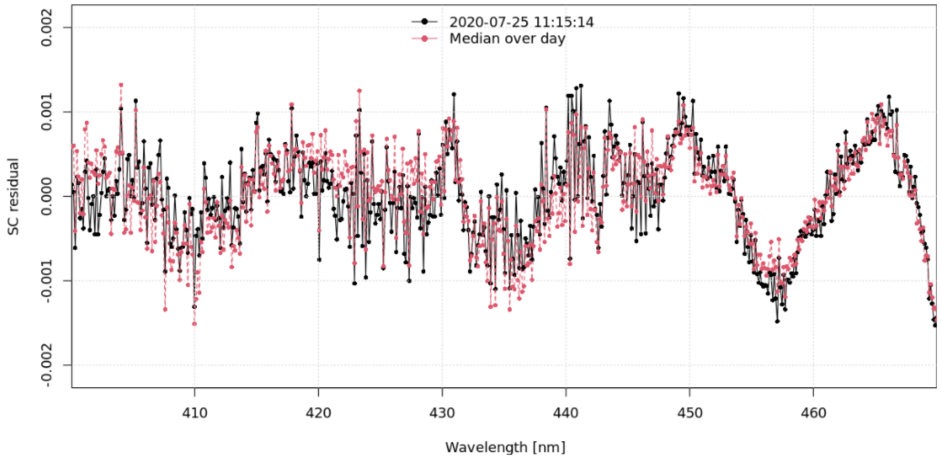


Spectral feature in the blue (LimBlueFeat) For some instruments, retrievals covering the blue spectral range show a distinct wave-like feature between about 450 and 470 nm. This feature appears to follow some seasonality, but other than that, does not show any clear correlation to a number of instrumental parameters. Hence it was so far not possibly to pin down the reason for this behavior.

An example is illustrated in Figure 105 where the spectral residuals (average over the day in red and one example timestamp in black) for one day in Rome-SAP for Pandora 117s1 is shown. The seasonality of this feature drives the weighted RMS of the spectral residual from 2e-4 when the feature is gone, to 6e-4 like shown in the figure, up to 1e-3 at its maximum.

Luckily, there is so far no evidence that e.g. NO_2 column retrievals (fitted between 400 to 470 nm.) are biased. However, again weak absorber in this spectral range, like CHOCHO, do suffer from this feature.

Figure 105: Spectral residual at 2020-07-25 in Rome-SAP from Pandora 117s1. The near noon residual is shown in black and the daily average in red. The wave-like structure around 460 nm is evident and features a pronounces seasonality.



Filter sensitivity (LimSens) There is evidence that due to some yet unidentified shortcomings in the filter characterization, spectral residuals remain after L1 data correction for spectra using a diffuser. For cases where the chosen reference spectrum and the measurements are based on the same diffuser, this residual feature cancels out in the ratio applied to get L2 data. Once different diffusers have been used (only for 2S Pandora systems) or a reference taken with diffuser is applied to a measurement without diffuser (e.g. lunar retrievals), features remain to different extents.

Also here the reason is not fully understood yet, but might be related to the quality of the spectral parameterization of the filter transmission or/and even a systematic issue for the diffuser transmission being a function of the divergence of the input beam (what obviously is different for a lab FEL lamp and the sun).

Currently discontinued data product development or limited usability:

Table 15: Limiting issue(s) and level of impact (already product development or its usability) for certain data products. The identified issues (LimUVFilt, LimBlueFeat, LimSens) are explained in the text.

Data product	Limiting issue	Impact	Comment
All S2 DS	LimSens	Usability	Without this issue, the data availability would be doubled. Currently it is like for the S1 product.
All DM	LimSens	Usability	Reference spectrum from solar data can lead to biased data. A lunar reference would need to be compiled.
Spectral AOD	LimSens, LimBlueFeat	Development	Absolute irradiance could be biased and additional wave-like features in the blue spectral range.
DS NO_2 from S1	LimBlueFeat	Usability	Columns are not affected but WRMS shows seasonality what impacts reliability of quality flagging.
DS CHOCHO	LimBlueFeat	Development	Appears to be severely impacted.
DS HCHO	LimUVFilt	Development	Some data sets are affected and need to be flagged out.
DS HONO	LimUVFilt	Usability	Appears to be severely impacted.

6 New data products schedule

The suggested schedule for new or refined data products introduced in section 3 is given in table 106. Since basically all new products include new concepts and ideas some of them even need modifications on the instrument to be done (e.g. the wedged window needed for weak absorber and an improved FOV to be mandatory for spectral AOD retrievals), delays in the development can always happen and will be communicated to ESA in time.

Each new or refined algorithm will be implemented following the implementation strategy outline in section 2.3.3 of [1]. The product release dates suggested in column 2 of Table 106 refer to the date, when the implementation phase is supposed to be completed. First and second stage for spectral AOD refer to section B.5. Products that have been suggested along the project, but could not be finalized due to the mentioned hardware limitations (see section 5), are listed in the last row of the table. Continuation on the development of those products is planned to be done within the follow up project FRM4AQ-2.

Figure 106: Suggested schedule for new data products within the project. Gray text refers to already completed data products.

Estimated release date		Data product(s)
PM2	July 2019	Surface concentration and tropospheric NO2 from sky data
PM6	June 2020	Improved NO2 total columns
MTR	December 2020	Direct sun total O3/O3temp Direct sun total SO2 Direct sun total HCHO Surface concentration and tropospheric HCHO from sky data HCHO profiles from sky data NO2 profiles from sky data
PM9	June 2021	Direct sun total O4 Direct sun total NO2 from spectrometer 2 Surface concentration and tropospheric NO2 from sky data from spectromter 2 Spectral AOD (first stage)
PM11	December 2021	Direct sun total O2 Direct sun total H2O Direct moon total NO2 Surface concentration and tropospheric column H2O Spectral AOD (second stage)
PM13	June 2022	Profiles for the same gases
FM	December 2022	Direct moon total columns for NO3
	Postponed to FRM4AQ-2	Direct sun total HONO Direct sun total CHOCHO

7 ATBD

This section covers deliverable D4, the ATBD of new/updated algorithms and TN & publications. The ATBD is given in in *Cede* [6], section 6. Here is an overview of the different algorithms included:

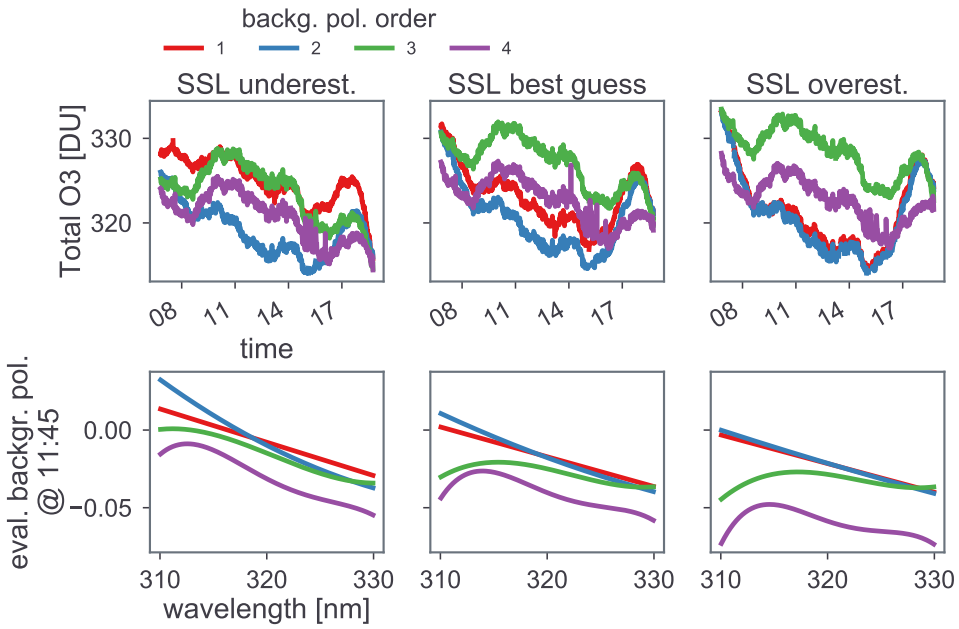
- Section 6.1, Alignment Algorithm: this gives the theoretical background and practical implementation of the algorithm, which calculates the tracker positions needed to point Pandora to a specific direction in the sky. It includes the question of co-alignment, i.e. how the software deals with Pandora-2S systems, which have 2 optical systems included.
- Section 6.2, Signal to Noise Ratio Optimization: this gives the theoretical background of how the software optimizes the signal to noise ratio (SNR) and discusses the cases of a single scan dark correction, multiple scans dark correction, use of the so-called dark map, and the inclusion of blind pixels.
- Section 6.3 - Data uncertainty: this section illustrates the concept of the uncertainty treatment within the BSS, including the definitions of independent (zero correlation length along the dimension of the independent error), structured (between zero and infinite correlation length) and common (infinite correlation length) uncertainties.
- Section 6.4, L1 Algorithm - Data Correction: this describes the Pandora raw data correction steps, i.e. Dark Correction, Non-Linearity Correction, Latency Correction, Flat Field Correction, Conversion to Count Rates, Temperature Correction, Stray Light Correction, Sensitivity correction and Wavelength Correction.
- Section 6.5, L2Fit Algorithm - Spectral Fitting: this describes the Pandora Spectral Fitting algorithm, starting with Lambert-Beer's law and covering topics such as convolution, Slant Optical Depth, the Basic Fitting Equation and its terms, the Least Squares Minimization for linear and non-linear fitting.
- Section 6.6, L2 Direct Algorithm: this describes the algorithm to produce the final data from the derived slant columns from direct sun measurements. Emphasis is set on procedures to refine the AMF calculations by taking into account effective temperature measurements or climatologies and/or slant column measurements or climatologies. Further data averaging under consideration of noise levels is addressed.
- Section 6.7, L2 Air-Ratio Sky Algorithm: the BSS trace gas profile algorithm is explained in this paragraph, which runs without radiative transfer calculations and a-priori profile information.
- Section 6.8, Climatologies: the real time nature of the BSS data levels requires climatological data for meteorological parameters like surface pressure and temperature. Further, gas effective (=vertical profile weighted) parameters like height and temperature and a stratospheric climatology of NO_2 help to constrain the retrieval and AMF estimation. All climatologies are introduced in this section and also how they are used to support uncertainty estimations.

A Implications when using a literature reference for total O_3 retrieval

One known source of a bias in total O_3 is when a constant O_3temp is assumed in the spectral fitting retrieval (see e.g. *Zhao et al.* [57]). At some occasions, comparisons to Brewer spectrophotometers and former comparisons to OMI [37] indicate systematic underestimation of the current standard total O_3 in the range of 2 %. In the following we try to find a possible explanation for these observations.

Due to the complexity of the issue we focus on one example day, measured at Huelva, Spain, on the 21st June 2019 by Pandora 121. The diurnal variation of total O_3 is shown in the top panels of Figure 107 for different situations: The individual colors refer to different background polynomial orders (BPol) used in the spectral fitting (1 to 4). The figure columns refer to different SpecSL correction scenarios with expected underestimation of SpecSL (left panel) to a best guess SpecSL correction (middle panel) and a expected overestimation (right panel).

Figure 107: Diurnal variation of total O_3 (top panels) for different selections of BPols (color-coding) and SpecSL correction scenarios (columns). The evaluated BPols at 11:45 are shown in the bottom panels. Explanations are given in the text.

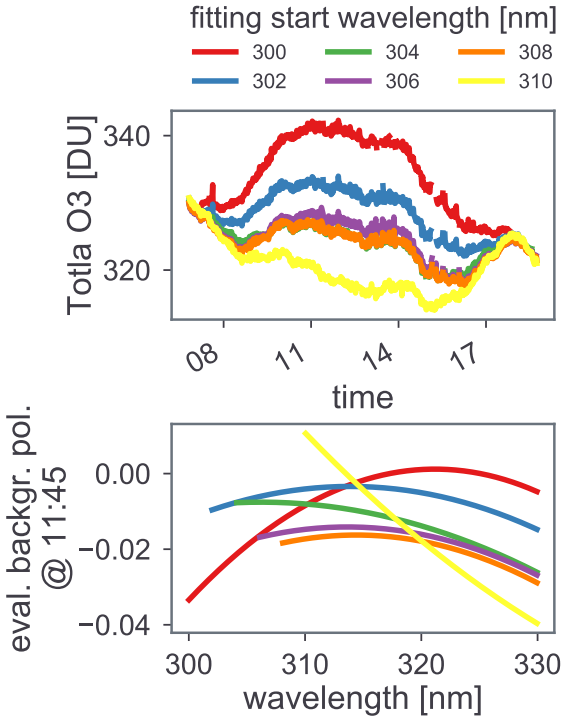


We can observe two notable things: First, there is a significant impact of the chosen BPol on the magnitude and diurnal variation of total O_3. Second, across the SpecSL correction scenarios, total O_3 patterns are rather constant for e.g. BPol 2 and 4, but more variable for BPol 1 and 3. In the bottom row of the figure the evaluated spectral BPols are shown. Looking at those variations across the SpecSL scenarios, evidently higher order BPols seem to scale with the SpecSL features. Also, the steeper the spectral decline from lower to higher wavelengths for a BPol, the lower total O_3 columns appear to get. In this cases BPol seems to "explain" parts of the O_3 absorption. The pretended advantage of BPols "correcting" SpecSL is actually a disadvantage, because it makes the SpecSL correction impact unpredictable and, as seen in the figures, behaves differently at different AMFs.

For the moment we accept this strong dependency on BPol and change now the view to look at the dependency of O_3 columns on the fitting window starting wavelengths ($\lambda 1$). The resulting O_3 columns for $\lambda 1$ ranging from 300 to 310 nm

are summarized in figure 108. The shown scenario is for BPol=2.

Figure 108: Diurnal variations of total O_3 (top panel) for different selections of $\lambda 1$ (color-coding). The evaluated BPols are shown in the bottom panel for 11:45.



Again, a huge impact on the diurnal pattern of total O_3 is discernible, with decreasing total O_3 amounts for increasing $\lambda 1$. The data differ by a SC bias, what is indicated by almost no difference on high AMFs.

So far we can summarize two main observations:

1. Total O_3 strongly depends on the selection of the BPol. Further, evaluated BPols tend to decrease for smaller wavelengths at higher BPols.

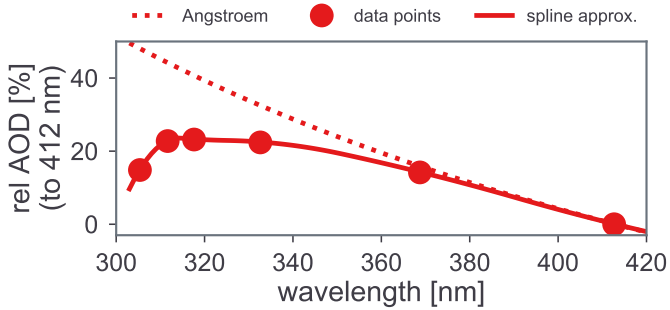
2. The shorter $\lambda 1$ the higher the O_3 columns get.

Ad 1.: Dependency of BPols on total O_3

For absolutely calibrated instruments when utilizing a literature reference, the evaluated BPol should explain the spectral attenuation of molecular scattering and aerosol extinction. In the case of the BSS algorithms, where molecular scattering is removed from the spectra before fitting, only aerosol extinction should in theory be left. If we assume that this is indeed the case for the measurements, it is notable that, in contrary to expectations (from Angstrom's law), the polynomial tends to go down again for lower wavelengths.

It has been reported already several times that the Angstrom approach should be extended by a curvature term [e.g. 41]. A study by *Carlund et al.* [19] however even suggested a decline in the spectral AOD for wavelengths below about 312 nm, but conceded a rather high uncertainty in this estimate driven by O_3 absorption. Figure 109 shows the suggested spectral AOD from this study (red dots), compared to a regular Angstroem approximation (dotted line). If indeed the spectral AOD is more complex in the UV, this might be an explanation for the strong impact of different BPols.

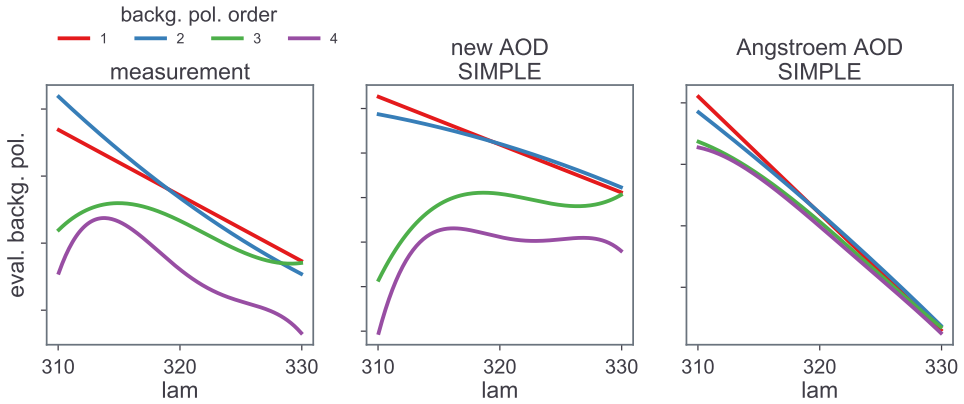
Figure 109: Spectral AOD as measured as campaign average in *Carlund et al.* [19] (red dots). Datapoints are given as percentage change relative to AOD at 412 nm. The smoothed spline extrapolated version of the datapoints (solid red line) is used as AOD parameterization in SIMPLE. For comparison, the commonly used Angstroem parameterization is shown as dotted line.



To test this we included this spectral AOD estimation (parameterized by the shown spline approximation) in SIMPLE to make a relative consistency check between the evaluated BPols. In the comparison plot in figure 110, we show on the one hand the already known results from the measurements, but also the retrieved BPols from SIMPLE simulations when using the new AOD parameterization from

the study (middle panel) or the classical Angstrom approach (right panel). Interestingly enough, using the new spectral AOD estimation the accordance between simulation and measurements is higher. These findings support the observations of *Carlund et al.* [19]. Further, as expected (but not shown here), the choice of BPol has less impact on the O_3 columns for the simulation using the Angstroem approximation.

Figure 110: Retrieved evaluated BPols from measurement (left panel) and simulation, using the new AOD parameterization (middle panel) and the Angstroem approach (right panel). The color-coding represents different BPols. When using the new AOD paramterization, the accordance between measurements and simulation is significantly higher.

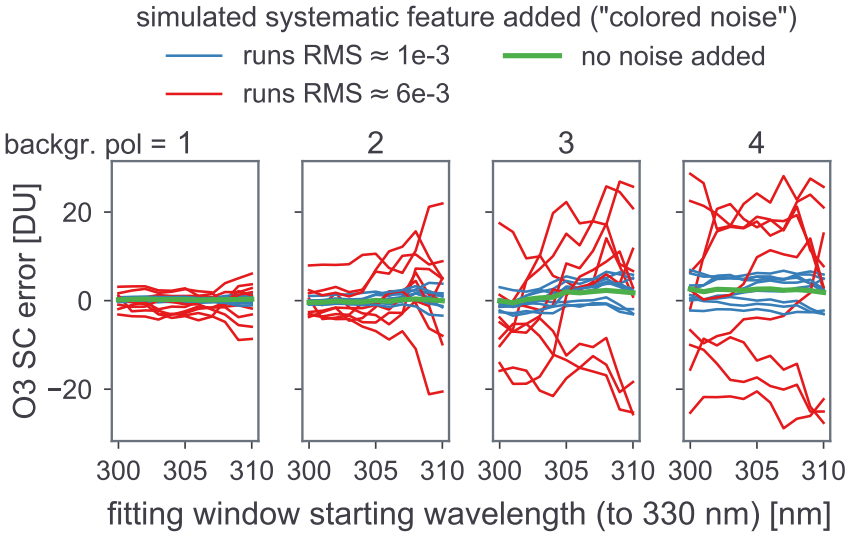


Ad 2.: Dependency on $\lambda 1$

Although Pandora instrument characterization is state of the art, a residual spectral structure in the order of RMS=6e-3 always remains in the spectral fitting (for the standard fitting window 310 to 330 nm) using the literature spectrum. Since this feature is highly repetitive across the Pandoras we hypothesize that if there is an impact of this feature it should impact all Pandoras similarly.

To probe a possible impact of this residual structure on the total O_3 data, we included a systematic "noise" feature generator in SIMPLE. Due to the fact that the exact representation of this residual structure is not possible (otherwise we would have corrected it already), we simulate this pattern by a Gaussian distributed noise with a power law spectrum [50]. By this, a so-called "colored-noise" spectrum can be created which exhibits, in contrast to a white noise spectrum, a certain dependency on the power spectrum (or in other words enhanced systematic spectral structures). Of course, it is very unlikely that by this approach the exact residual structure can be meet, but a generalized analysis of a possible impact is still valid.

Figure 111: Simulated SC error for retrievals using a literature spectrum as reference. A "colored-noise" pattern (spectral feature) is added to the measurements to simulated the spectral residual what are seen for measurements. Figure columns refer to different BPols and the color-coding refers to scenarios yielding about RMS 6e-3 (red) and 1e-3 (blue). As reference the noise free case is shown in green. Explanations are given in the text.



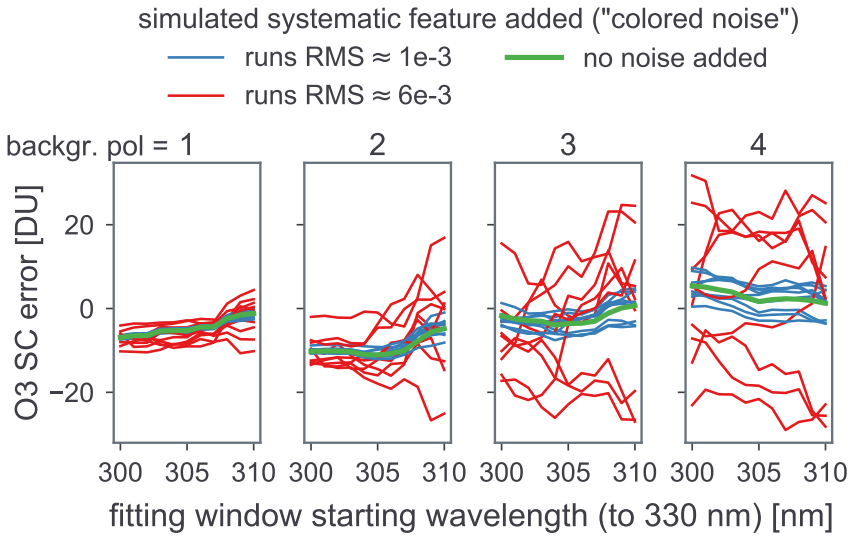
The simulated effect of such residual structures on the O_3 SC is shown in Figure 111 for different BPols (panels from left to right). To get an impression about the "distribution" of the effect we show the results for 10 different runs and two different RMS levels: 6e-3 (as seen in the data) in red and 1e-3 in blue. For reference, the scenario without additional residual structure is shown in green. We can draw several conclusions from this plot.

- Systematic structures can lead to over- and underestimation of O_3 columns.
- The impact of systematic structures with a RMS in the order of 1e-3 is strongly reduced.

- For low order BPols, lower $\lambda 1$ seem to stabilize the retrieval, because the O_3 absorption features are stronger in the low wavelength range.
- For higher order BPols, even lower $\lambda 1$ show differences.

For completeness, we also did this analysis when using the new spectral AOD parameterization as explained in the previous paragraph (figure 112). The general pattern stay the same, but now it can be seen that low order BPol are not able to get the complex AOD structure and hence are more prone to lead to underestimation.

Figure 112: Same figure as 111, but the simulation was done with the new AOD parameterization as shown earlier.



Remarks when using a literature reference spectrum for total O_3 retrieval

1. A strong impact of the selection of BPols might be connected to a more complex structure of spectral AOD (as shown). A higher order BPol would be needed to represent this accordingly.
2. BPols > 1 are also capable to explain residual SpecSL features (not only the so-called offset polynomial), which might be not always wanted. In this case the evaluated BPol would not represent AOD extinction.
3. BPols = 1 seems to be able to can explain partly O_3 absorption. This might be due to the fact that its shape is far off explaining this more complex spectral AOD and might "fit" better to the O_3 background.
4. Evaluated BPols are inclined to decrease for smaller wavelengths and higher BPols, supporting a more complex spectral AOD variation.
5. Residual structures due the usage of a literature reference can in principle cause systematic biases in the O_3 data.
6. The probability for that to happen starts with an RMS larger ≈ 1e-3.
7. The probability is lowest for low order BPols and shorter starting wavelengths. But this setup relates to point 3 which leads to systematic underestimation.
8. The probability is highest for higher order BPols and the probability for an improvement for shorter starting wavelengths is reduced.

The implementation of the reported observations in SIMPLE is key in closing gaps between model and measurements, which finally is crucial for the in depth error estimation regarding the new O_3 product (see section 3.5.2). By considering these remarks, the Pandora total O_3 product using a literature reference ("out of the box" O_3) can be significantly improved (e.g using a BPol 2 or 3 and using lower starting wavelengths).

We argue that the conclusions from above are not limited to Pandoras, but apply to all O_3 algorithms applying a literature reference spectra.

B About spectral AOD from Pandora

B.1 Introduction

This appendix is an overview of how to derive spectral AOD from Pandora direct sun observations and what sources are contributing to the total uncertainty in it. The idea of this study is to rank the different uncertainty contributions in order to identify the issues, which should be worked on first to reduce the overall uncertainty. All data given here are estimations and may need to be re-evaluated in a more rigorous analysis. Note that all uncertainties in this appendix are expressed at the "2σ level" (2σLEV), which corresponds to a 95% confidence interval when assuming a normal distribution. This must be considered when comparing them to literature values, where most often the "1σ level" (1σLEV) is cited. To emphasize this, we use for the uncertainty the acronym 2σUNC here.

B.2 Basic equation

The basic equation to derive spectral AOD is given in equation 80 in *Cede* [6], here expressed in a slightly different way:

$$F = F_0 \cdot \exp\left(-m_{AER} \cdot \tau_{AER} - m_{SCA} \cdot \tau_{SCA} - \sum_i m_{GASi} \cdot \tau_{GASi}\right) \quad (10)$$

Equation 10 is applied to each single wavelength. Although each parameter in equation 10 is wavelength dependent (some more, some less), we do not explicitly write this here.

F	Signal at the measurement location corrected to the standard Sun-Earth distance
F_0	Signal outside the atmosphere at the standard Sun-Earth distance
m_{AER}	Aerosol (direct sun) AMF
m_{SCA}	Molecular scattering AMF
m_{GASi}	Gas extinction AMF for gas i
τ_{AER}	AOD; that's what we are looking for
τ_{SCA}	Molecular scattering optical depth (OD)
τ_{GASi}	OD for gas i

Solving equation 10 for τ_{AER} we get

$$\tau_{AER} = \frac{1}{m_{AER}} \cdot \left(\ln F_0 - \ln F - m_{SCA} \cdot \tau_{SCA} - \sum_i m_{GASi} \cdot \tau_{GASi}\right) \quad (11)$$

Hence the "AOD retrieval algorithm" is a one-liner. The difficulty in deriving spectral AOD is not at all the algorithm itself, but rather the knowledge and uncertainty of all the input parameters at the right side of equation 11.

B.3 Description and uncertainty estimation for each parameter in equation 11

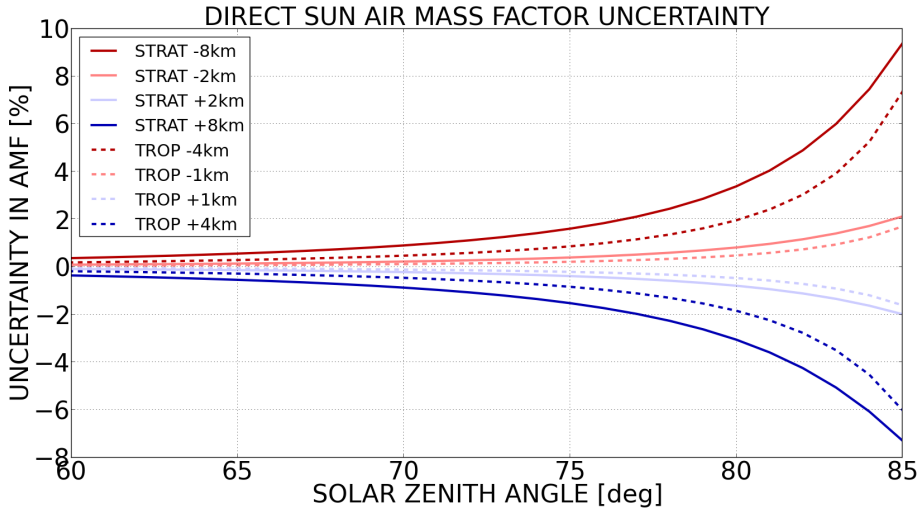
In this section the uncertainty for each parameter at the right side of equation 11 is discussed.

B.3.1 Air mass factors

The AMFs m_{AER}, m_{SCA} and m_{GASi} are calculated for direct sun geometry with a simple formula (equation 118 of *Cede* [6]), which is a function of the (refraction corrected) solar zenith angle (SZA) and the effective height (HEFF) of the species. The wavelength dependence in the direct sun AMF can be neglected. The error in the AMF calculation can be estimated by varying HEFF in the equation. Figure

113 shows the error in the AMF caused by a different HEFF as a function of SZA. It is shown for a stratospheric gas at a nominal HEFF of 22.4km (labeled STRAT, e.g. O_3) and a tropospheric gas at a nominal HEFF of 4.2km (labeled TROP, e.g. HCHO). Values for m_{AER} and m_{SCA} would be similar to the tropospheric gas. The light red and blue lines are for typical variations in HEFF (approximately 1σLEV) and the dark red and blue lines for extreme variations (approximately 2σLEV). One can say that the 2σUNC in the AMF is 0.0%, 0.4%, 0.6% and 2.5% for SZA 0°, 60°, 70° and 80° respectively.

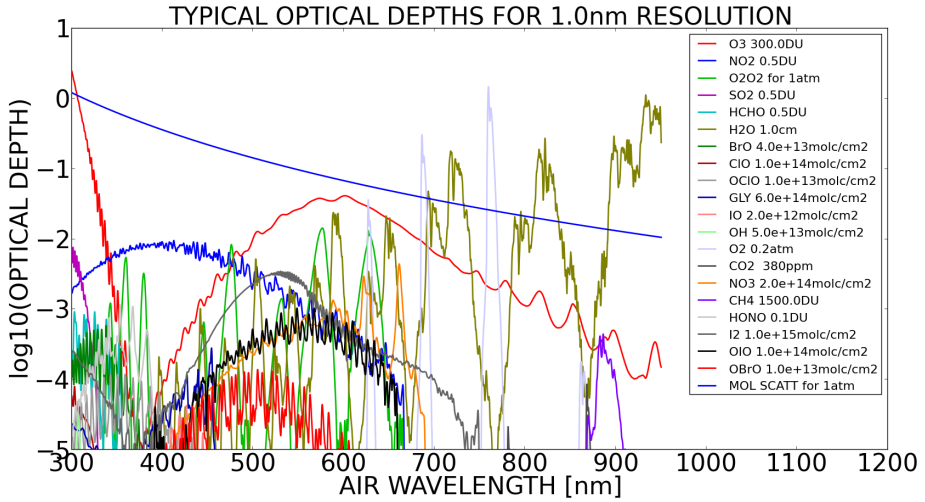
Figure 113: Estimated uncertainty in direct AMS as a function of SZA



B.3.2 Molecular scattering optical depth

The molecular scattering OD τ_{SCA} at standard atmospheric pressure (1013.25hPa) is shown in figure 114 (labeled MOL SCATT) and is also given in the instrument calibration file as a function of wavelength. It is then corrected for the typical station pressure using a climatological function depending on the station altitude. The difference between the actual station pressure at the ground location and the typical pressure is the main uncertainty source for τ_{SCA}. Pressure typically varies about 2% in time (2σLEV) around the climatological mean, which directly translates to 2σUNC=2% for τ_{SCA}. If a pressure sensor would be available (e.g. included in the Pandora spectrometer system), the uncertainty in τ_{SCA} would go down significantly to 2σUNC=0.02%.

Figure 114: Optical depths for "typical" columns amount of trace gases in the Pandora database and for molecular scattering (MOL SCATT) as a function of wavelength.



B.3.3 Gas optical depth

Figure 114 shows typical ODs of the gases included in the Blick Software Suite database in a logarithmic scale over the Pandora wavelength range. The gas name as used in *Cede* [6] and the "standard" column amount are given in the figure legend. As can be seen, τ_{GAS} is very small for many of the gases. While gas retrievals using spectral fitting technique can still be successful for OD>1e-4 and possibly even for OD>1e-5, the influence of these weak absorbers on the spectral AOD retrieval is very small.

Figure 115: Estimated uncertainty in ODs of trace gases as a function of wavelength.

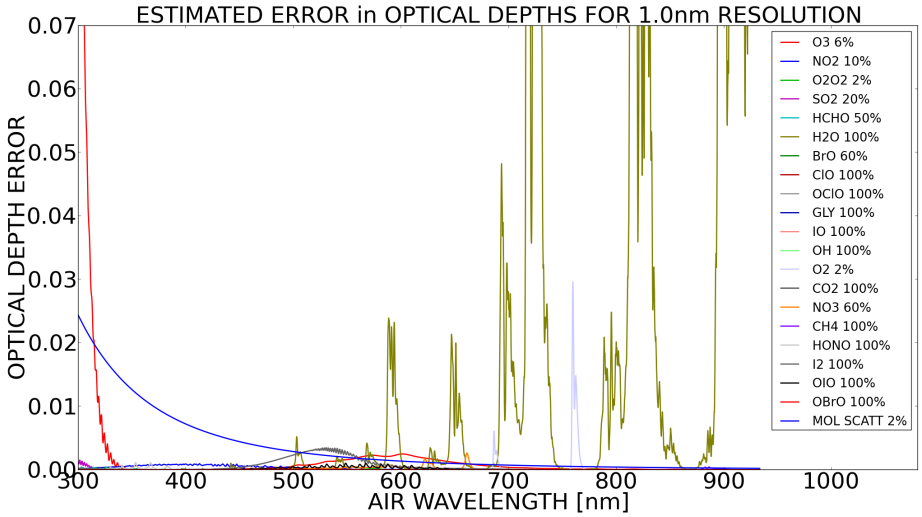


Figure 115 shows the estimated 2σUNC in the gas ODs in a linear scale, which is based on the column amount from figure 114 and the percentage retrieval 2σUNC estimations given in the legend of figure 115. The reason for the small error in the ODs for O_2O_2 and O_2 is not, that we can retrieve them so accurately, but rather that their values can be calculated using the surface pressure just as for molecular scattering. One can say that the gases not "seen" in figure 115 have no significant influence on the spectral AOD retrievals. This leaves basically O_3, O_2, molecular scattering and above all H_2O as the gases to influence spectral AOD over the ultraviolet (UV) and visible wavelength range. Note that NO_2 is not seen in figure 3, since we can measure it rather accurately. If we would not include it in the algorithm, then its influence would be significant for wavelengths below 550 nm.

B.3.4 Signal outside the atmosphere

F_0 is the Pandora L1 data, which would be measured outside the atmosphere at the standard Sun-Earth distance. So far our attempts to perform absolute radiometric calibration in the laboratory and transfer this into the field without the field calibration tool were not really successful, which means we have to rely on field calibration techniques to obtain estimations of F_0. A previous study [32] has shown that even for non-pristine sites we may achieve a 2σUNC of 5% in F_0 from Langley calibration. The quality of the Langley determined F_0 depends on:

- the stability of the total OD over the day. The more OD varies, the larger is the error in F_0.
- the magnitude of other SZA-dependent effects influencing the data (e.g. stray light described in the section B.3.5).

In the same study we have seen yearly drifts of F_0 on the order of 10% (2σLEV). We believe that this uncertainty can be reduced to 2% by applying regular calibration on an instrument, such as repeated Langley extrapolations and/or visits with reference units and a field calibration tool.

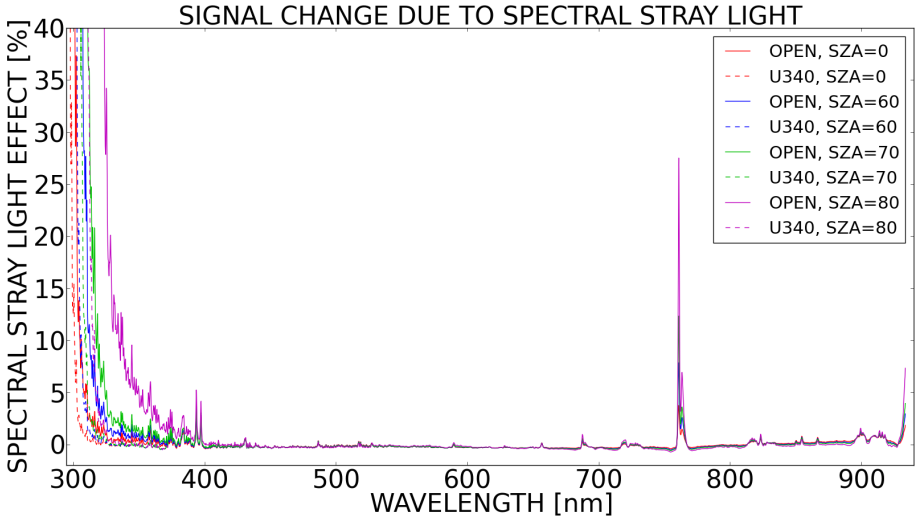
B.3.5 Measured signal

F is the Pandora L1 data corrected to the standard Sun-Earth distance. The correction to the standard Sun-Earth distance has negligible uncertainty. F suffers from long term (>one day) effects (drift of the radiometric sensitivity) and short term (<one day) effects such as

- Spectral stray light
- Spatial stray light
- Pointing uncertainty leading to sensitivity variations due to the non-perfect FOV
- Sensitivity variations due to different fiber bending

Uncertainty caused by long term drift has been attributed to the signal outside the atmosphere F_0 (section B.3.4) and is therefore not discussed again here. The short term effects are described in following paragraphs.

Figure 116: Simulated signal change due to spectral stray light for different SZAs as a function of wavelength.



Spectral stray light As a single monochromator, Pandora "suffers" from SpecSL. This means that the measured signal is changed compared to the signal a stray-light-free instrument would measure. It is basically a redistribution of the signal from wavelength regions of high intensity to wavelength regions of low intensity. Hence maxima in the spectra are reduced by SpecSL and minima are enhanced. The magnitude of the SpecSL depends on:

- the SpecSL-characteristics of the instrument.
- the "structure" of the incoming light. E.g. the SpecSL-effect for a flat input (same signal in all wavelengths) is much smaller than for a highly variable input.

Figure 116 shows the simulated direct sun signal change due to SpecSL for different SZAs as a function of wavelength in percent. The simulations are done for the characteristics of Pandora 110. The solid lines refer to OPEN in the filterwheel (spectrometer 1 for wavelengths below 527 nm and spectrometer 2 above 527 nm). The dashed lines are for spectrometer 1 with U340 in place. The relative SpecSL effect is most pronounced for regions of very low signal compared to the rest of the spectrum, e.g. in the low ultraviolet (UV) range for high SZA or in the oxygen-α-band around 760 nm. As expected, the influence of SpecSL is smaller, when U340 is used.

While the lines in figure 116 represent the error in F, the effect on the retrieved AOD caused by SpecSL cannot directly be derived from them, since it depends on the "interaction" between F and F_0. Imagine that the relative signal change due to SpecSL was constant with SZA (as it is approximately the case in some wavelength regions, e.g. 500-700 nm). In this case also F_0 would have the same enhancement, if it is obtained by Langley extrapolation methods. Consequently F and F_0 would have the same relative error, which would cancel out in equation E2 and would not affect the AOD retrieval at all. Unfortunately the SpecSL is not independent of SZA, especially in the UV (see figure 116), which causes a wrong value for F_0. How large this error is depends on the exact way the Langley calibration was done, e.g. the selection of SZAs.

For this study we simply assume a 2σUNC of the values in figure 4 in case no SpecSL-correction is done in the L1 processing and a reduction of the values in figure 4 by a factor of 5 in case a SpecSL-correction is applied.

Spatial stray light In direct sun observation mode, a portion of the circumsolar light is entering the instrument together with the direct beam. This additional light, which we call spatial stray light (SpatSL), enhances the "pure direct sun" signal. The enhancement depends on:

- the size of the instrument's field of view (FOV), in our case of circular shape with 2.5° FWHM.
- the amount of forward scattered light, which increases with SZA and aerosol loading.

Figure 117 shows the simulated direct sun signal enhancement due to SpatSL for different SZAs as a function of wavelength in percent. It is calculated for continental average aerosols with an AOD of 0.05 at 1000 nm.

This signal enhancement is not included in equation 10 and therefore introduces an error in the AOD retrieval. As for SpecSL, the lines in figure 117 represent the error in F, but cannot be directly translated into an error in the retrieved AOD for the same reasons explained in the previous paragraph. How large the error is depends

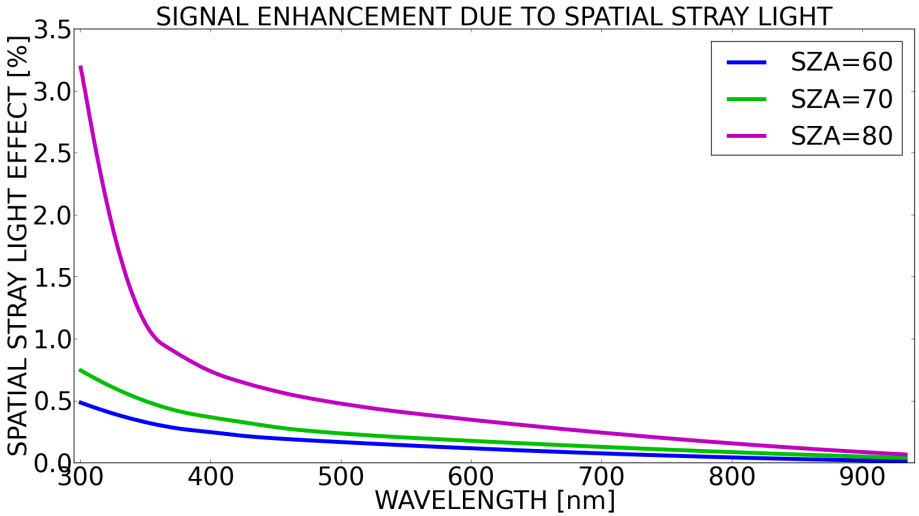
on the exact way the Langley calibration was done, especially the aerosol loading in the atmosphere during the day(s) used for the Langley calibration.

Here we assume that SpatSL is not corrected in the L1 processing and that the 2σUNC caused by it is half of the values in figure 117.

Pointing effect Despite using the internal diffuser for direct sun observations, the top part of the Pandora FOV is still not really flat for most instruments, which is a major issue in the measurements. Despite the excellent pointing accuracy of 0.1° for Pandora, this issue still gives us an estimated 2σUNC of 6%.

Fiber bending In previous studies [32] we have seen signal variations over the day, which we could attribute to the fiber bending. This issue was reduced by adding the fiber guide to the system. We estimate the 2σUNC due to the fiber bending to 4% without fiber guide and 1% with fiber guide.

Figure 117: Simulated signal enhancement due to spatial stray light for different SZAs as a function of wavelength.



B.4 Overall uncertainty estimation

B.4.1 Present situation

Figures 118 and 119 show the estimated error in the retrieved spectral AOD based on the assumptions outlined in section B.3 at SZA=0° and 70° respectively as a function of wavelength. F0Cal stands for the uncertainty in the F_0 calibration and F0Drift for the yearly drift in F_0. Here we assume the situation after one year of measurements without recalibration. FOV is the pointing effect and FIB stands for the fiber bending effect without fiber guide. tO3 and tH2O stand for the effects of τ_{O3} and τ_{H2O} respectively.

The black line is the combined uncertainty, where all contributions are assumed to be uncorrelated, except for the pressure dependent parameters τ_{SCA}, $\tau_{O_2O_2}$, and τ_{O_2}, which are assumed fully correlated (in the legend listed as SCA, O_2, O_4). REST is the combination of all error sources described before, but not singled out in the legend. The dashed black line is the "target 2σUNC" of 0.04. This is based on the stated 1σUNC of 0.02 for AOD by AERONET.

Figure 118: Estimated error contributions for spectral AOD in the present situation at SZA=0°.

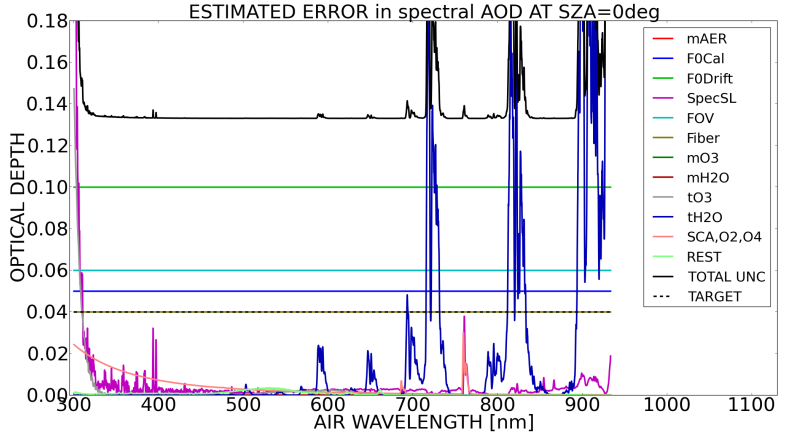


Figure 119: Estimated error contributions for spectral AOD in the present situation at SZA=70°.

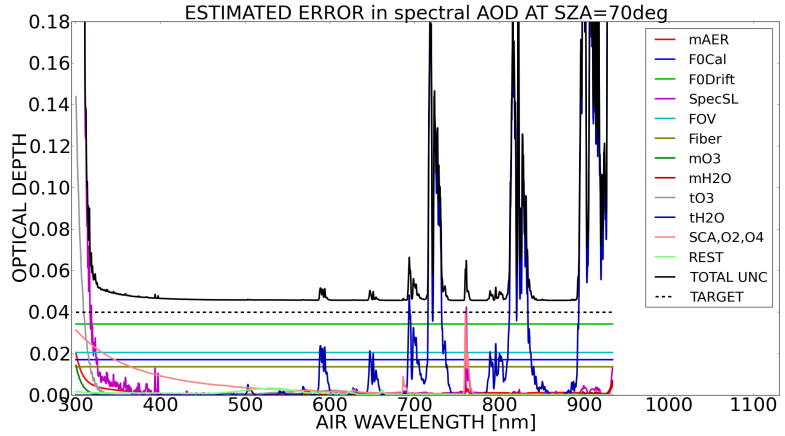
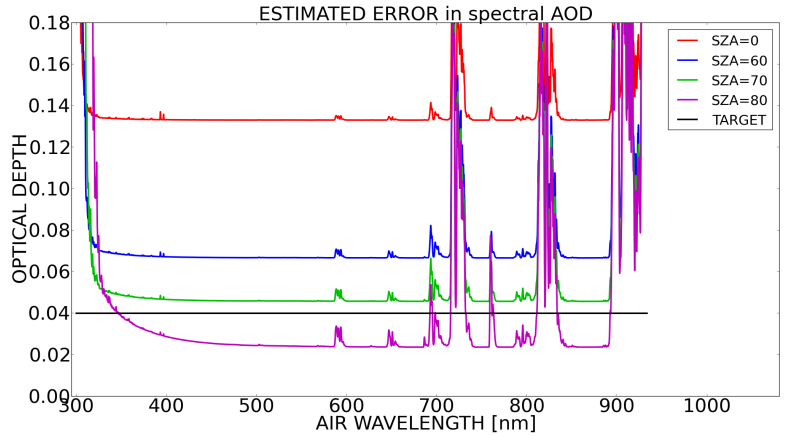


Figure 120: Estimated error in spectral AOD in the present situation at SZA=0°, 60°, 70° and 80°.



As observed in figures 118 and 119, several single error contributions prevent the goal of 0.04 to be reached: F0Drift, FOV, F0Cal, Fiber, tH2O, SpecSL and tO3:

- F0Drift: this uncertainty has to be reduced to 2% by the means outlined in section B.3.4 (applying regular calibration).
- FOV: also this uncertainty has to be reduced to 2%. At present we do not really know how to reach this goal.
- F0Cal: for this uncertainty we have to reach 2σUNC=2% too. That means just Langley calibration to obtain F_0 will not be enough on most network locations.
- Fiber: the Fiber Guide is mandatory to reduce this uncertainty to 1% as stated in section B.3.5.
- tH2O: this contribution can be reduced by developing an accurate water vapor retrieval for Pandora. Until this is done, one will need to stay away from spectral regions, which exceed the target uncertainty due to tH2O.
- tO3: we believe that with AXC we can reduce the 2σUNC for tO3 from 6% to 2%. In addition one will also have to stay away from the low UV range.
- SpecSL: this can be reduced by applying stray light correction, but still the low UV range will have to be avoided.

The estimated total uncertainty is shown in figure 120 for SZA 0°, 60°, 70° and 80°. It is obvious that in the present situation making an operational data product for AOD in Pandonia is not useful.

B.4.2 First stage improvement

We believe a meaningful spectral AOD data product can be made in Pandonia when ALL of these improvements have been made, which we call “first stage improvement”:

- F0Cal and F0Drift are improved using a full calibration procedure involving the use of stationary and mobile reference units, field calibration tools and automated application of Langley techniques.
- The shape of the FOV is improved, although we do not really know how to do this yet.
- Each instrument has a fiber guide.
- AXC is used to get accurate tO3.
- L1 data are corrected for spectral stray light.

The effect of these improvements is shown in figures 121 to 123.

Figure 121: Estimated error contributions for spectral AOD after the first stage improvement at SZA=0°.

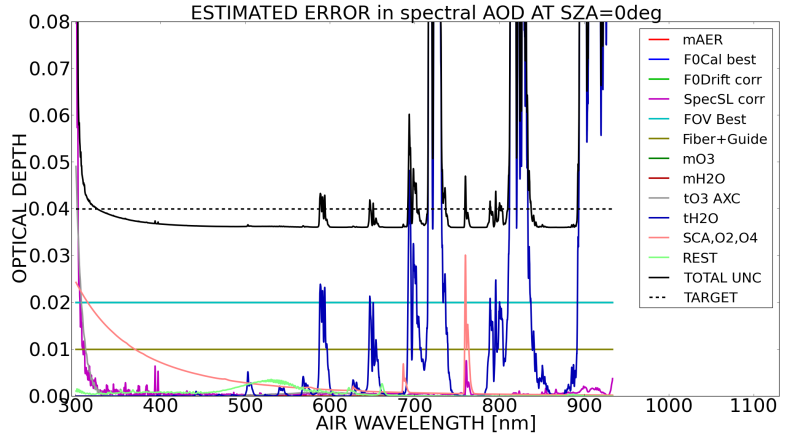


Figure 122: Estimated error contributions for spectral AOD after the first stage improvement at SZA=70°.

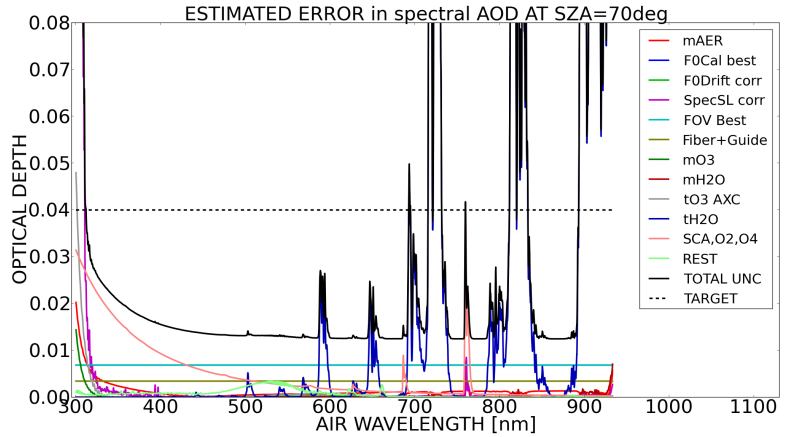
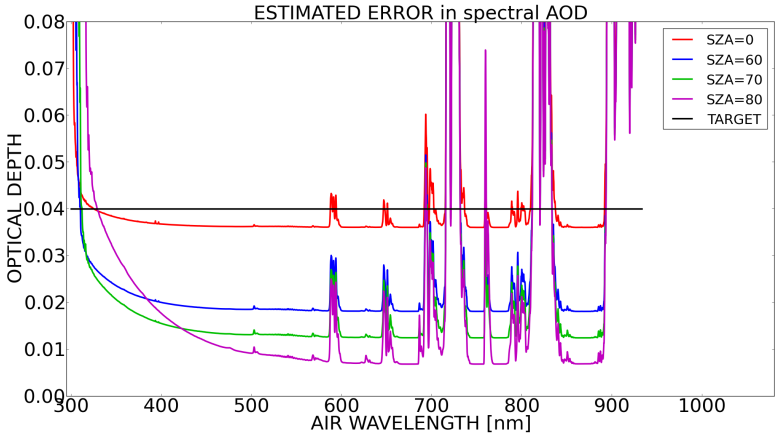


Figure 123: Estimated error in spectral AOD after the first stage improvement at SZA=0°, 60°, 70° and 80°.



As seen in figure 123, the spectral AOD after the first stage improvement is already meeting the target except for some wavelength regions, which are:

- Below 329 nm mostly due to SpecSL
- Between 758 and 761 nm due to SpecSL around the oxygen-α feature
- Around water vapor absorption regions, 587-595 nm, 646-651 nm, 691-703 nm, 715-736 nm, 788-837 nm and above 891 nm

B.4.3 Second stage improvement

In a so-called "second stage improvement" the following changes could be made:

- A pressure sensor is added to Pandora, which reduces the contribution of SCA,O_2,O_4 by a factor of 100.
- A water vapor retrieval is developed with a 2σUNC of 6% for tH2O.

The effect of these improvements is shown in figures 124 to 126. It would meet the target except for wavelengths below 320 nm (mostly due to SpecSL) and above 928 nm. Hence only after the second stage improvements one could claim that Pandora derives spectral AOD continuously between 320 and 928 nm.

Figure 124: Estimated error contributions for spectral AOD after the second stage improvement at SZA=0°.

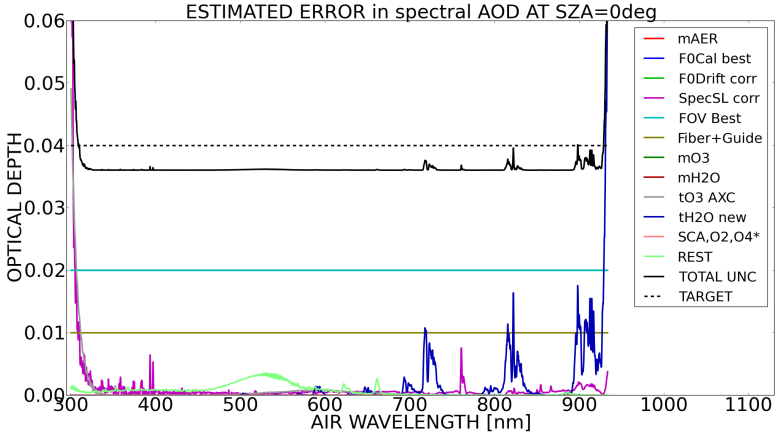


Figure 125: Estimated error contributions for spectral AOD after the secind stage improvement at SZA=70°.

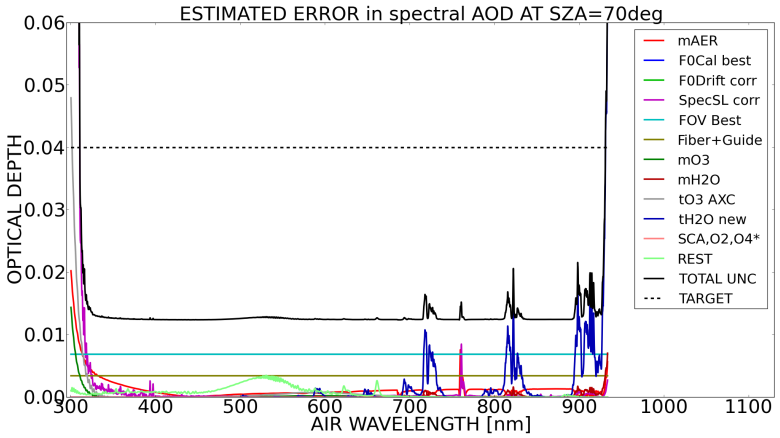
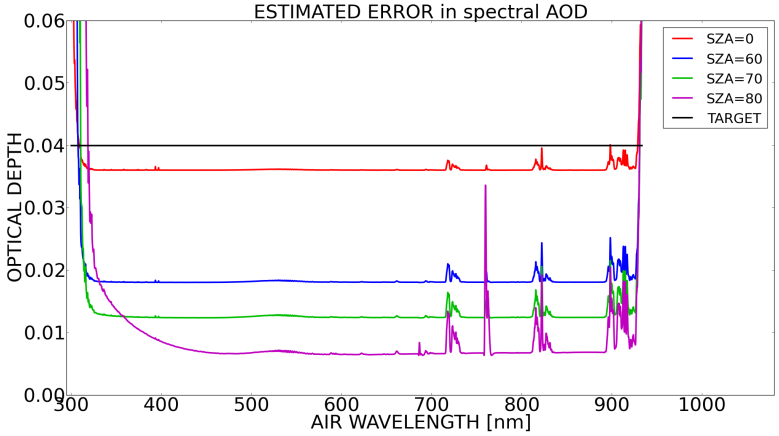


Figure 126: Estimated error in spectral AOD after the second stage improvement at SZA=0°, 60°, 70° and 80°.



B.5 Conclusion

LuftBlick is repeatedly asked, whether there is already an algorithm to derive spectral AOD from Pandora. Deriving AOD from direct sun observations is a one-liner (equation 11) and therefore the algorithm is not at all the problem. Instead the question should be, whether Pandora optical properties, L1 data correction steps and trace gas algorithms for strong absorbers are already in a shape to derive spectral AOD for Pandora.

The answer to this question is: In the present situation retrieving spectral AOD from Pandora is NOT useful due to several issues pointed out in this study. The following improvements need to be made at the minimum in order to produce meaningful spectral AOD ("first stage improvement"):

- Full calibration procedure involving the use of stationary and mobile reference units, field calibration tools and automated application of Langley techniques have to be applied to get proper values of F_0 at any time.
- The FOV must be improved. The subject needs to be a major focus over the next months.
- Each instrument needs to have a fiber guide.

- AXC needs to be applied to derive the total ozone column.
- L1 data need to be corrected for spectral stray light.

After the first stage improvement, Pandora spectral AOD can be retrieved between 329 and 891 nm except for some water vapor absorption regions, where the uncertainty still exceeds the allowed maximum of 0.04.

One should also consider the following second stage improvement:

- A pressure sensor is added to Pandora.
- A water vapor retrieval is developed with a 2σUNC of 6% for total water vapor columns.

After the second stage improvement, Pandora spectral AOD can be continuously retrieved between 320 and 928 nm.

C Note on resolution change sensitivities for trace gas retrievals

The accurate determination of the instrument slit function (or response function) is crucial for optical spectroscopy (e.g. for DOAS applications). Good characterization in the laboratory is one thing, but maintaining it during field operation is usually a different story. If the slit function is applied in a parameterized way, analytical methods exist to keep track and sufficiently correct changes of the slit function. This is usually realized by adjusting the so-called resolution parameter (e.g. the FWHM of the slit function). This is e.g. done in the BSS and formalized in a general way by *Beirle et al.* [15].

Beside the intrinsic temperature dependency of the slit function (i.e. the thermal response of the optical bench), a change in the exact illumination of the spectrometer might also change the spectral dependency of the slit function. This is thinkable e.g. by a change in the fiber alignment, in particular if a lens is part of the optical system (as for the Pandora).

Strong and unstructured absorber like O_3 and NO_2 are expected to be affected by this to a lesser extend, but the retrieval of weakly absorbing gases (like HCHO) might be significantly impacted. To quantify this effect, we conducted a series of simulations (with SIMPLE) and evaluated some field data.

It needs to be noted that the impact of a resolution error is dependent on which "spectrum" is affected: we can distinguish three scenarios:

- The resolution has changed e.g. due to transport to the site, but stays stable in the field: in this scenario, both the measurements and the Frauhofer reference (selected from this site) agree in terms of resolution, but the convolution of the high resolution cross sections is not valid anymore. We abbreviate this scenario with **ResConv**.
- The resolution is correct for a certain time - including the time when the Fraunhofer reference was picked - and changes afterwards. In this scenario the Fraunhofer structures are not canceled out sufficiently anymore. We abbreviate this scenario with **ResFraun**.
- Most likely, we will encounter a combination of scenario ResConv and ResFraun: **ResConvFraun**.

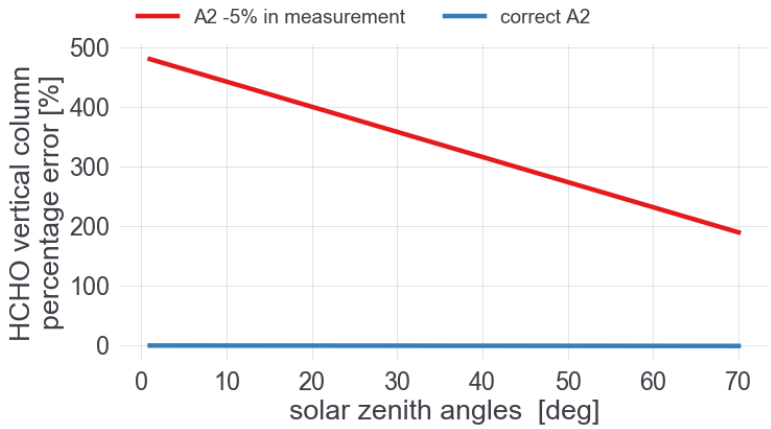
Due to the nature of the MAX DOAS algorithms (using sequential Fraunhofer reference spectra), only scenario ResConv could potentially play a role for these kind of retrievals.

C.1 Impact on direct sun total HCHO

Highest interest in this respect was set to HCHO, since, as reported e.g. in section 3.7, some discrepancies observed in the data require still more explanation.

The simulations clearly revealed a striking resolution impact of about 100 % HCHO VC error per 1 % resolution change for scenario ResFraun (a negative resolution error leads to positive VC error). This is exemplarily shown for two SZAs (1 and 70 deg) in Figure 127 (data with biased resolution in red), where it can be seen that the impact is highest for smallest SZAs. Consequently a resolution error of this type is expected to introduce an artificial SZAs dependent variation. Note that the figure legend talks about parameter *A2*, which is tantamount to *resolution*.

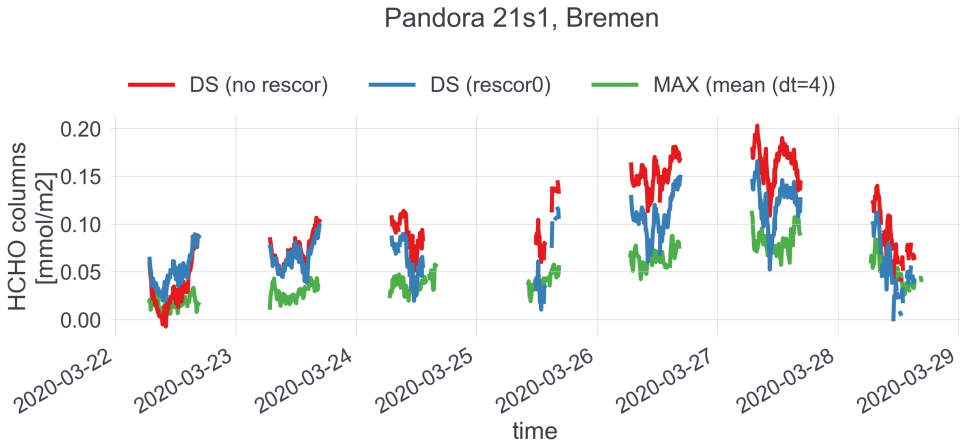
Figure 127: Simulated resolution (or A2) sensitivity on the direct sun total HCHO VCs for scenario ResFraun (explanations in the text). The red curve corresponds to a -5% resolution change in the measurements. The blue curve is for a correct resolution.



Scenario ResConv instead, yields a substantially smaller impact of about 1 % VC error for a 1 % resolution error (this time a positive resolution error stimulates a positive VC error).

As mentioned, the BSS is able to account for resolution changes (both scenarios ResConv and ResFraun) in the spectral fitting, which could be also verified based on synthetic data (not shown here).

Figure 128: Time series of total (DS) and tropospheric (MAX) HCHO columns for Pandora 21s1 in Bremen for one week end of March in 2020. Data points in red refer to the standard HCHO direct sun product and in blue when a constant resolution change is fitted as well. Data points in green show the MAX data from the same instrument.



A test application is shown for Pandora 21s1 (Bremen, Germany) in Figure 128. The comparison of HCHO tropospheric columns from Pandora MAX data (in green) and standard direct sun data (in red) reveals a clearly changing bias over several days. This changing bias is strongly removed with activated resolution change fitting (illustrated in blue). The resolution corrected data change on average up to 30 % which is in line with the retrieved resolution change of (only!) 0.3 % and comes along with an overall reduction of the fitting RMS.

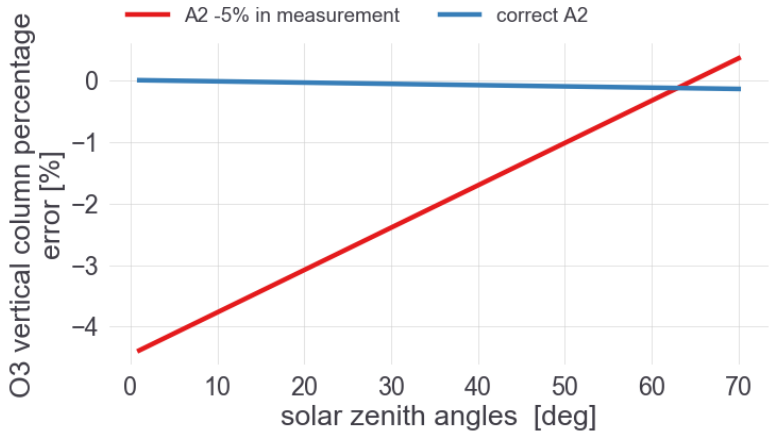
As a consequence of this results, it is strongly recommended to active resolution fitting for HCHO retrievals. This is already routinely done for a number of satellite missions, but so far not for Pandora retrievals.

Please note that the combined scenario ResConvFraun is not corrected in the example. This would require the resolution correction of the measured Fraunhofer reference spectrum as part of the field calibration procedure.

C.2 Impact on direct sun total O_3

The resolution sensitivity for the direct sun total O_3 product - based on a measured (=synthetic) reference spectrum - is very small compared to HCHO, but not negligible. As can be seen from Figure 129, a 1 % resolution error leads to a $\approx$ 1 % error in the retrieved O_3 temperature, which translates also to a $\approx$ 1 % error in the VCs. Since a rather strict accuracy requirement is set to nowadays reference O_3 datasets, resolution changes need to be accounted for.

Figure 129: Simulated resolution (or A2) sensitivity on the direct sun total HCHO VCs for scenario ResFraun (explanations in the text). Same definitions as for Figure 127.



Note that the resolution chance fitting cannot be activated for the O_3 product (or any other product) using a literature reference, because the retrieved VCs are artificially biased as seen from simulations (and measurements).

C.3 Impact on direct sun total NO_2 and SO_2

Tests on synthetic and measured data compliantly confirm a negligible effect of resolution changes on direct sun total NO_2 and SO_2, although a positive impact on the fitting RMS can be seen. Hence also for this data products resolution change fitting should be activated.

C.4 Summary and conclusions

We could identify a significant impact of resolution changes on trace gas retrievals, in particular on weakly absorbing gases. Although Pandora is considered to have a very robust slit function, resolution sensitivities of 100 % VC error per 1 % resolution error for e.g. HCHO are striking. It is very likely that other weakly absorbing gases (e.g. HONO, BrO, ...) are affected in a similar way.

Although the impact is rather small for O_3 retrievals, it is still significant in view of envisioned accuracies of below 1 %. As could be expected, NO_2, and luckily SO_2, are invariant in terms of resolution changes.

Due to the overall very small impact of scenario ResConv, all MAX DOAS products are basically not affected by resolution changes.

As a consequence of the observed resolution sensitivity we have

- activated the resolution fitting for the operational data products of NO_2, SO_2, O_3 and HCHO.
- started to upgrade the BSS to allow the consideration of resolution changes also for the (synthetic) reference spectrum as part of the field calibration.
- started to investigate the sensitivity of other slit function parameters next to the resolution (= "width" of the slit function), e.g. the shape of the slit function.

D Statistical framework for dataset intercomparison

D.1 Introduction

In the following, we introduce a framework for comparing co-located Pandoras as a first stage (InterComp). In principle, the approach can, and will be adjusted to compare different datasets of Pandoras (CrossComp), and also to compare different datasets in general that are supposed to measure the same, e.g., external validation to Brewers for O_3. To illustrate the approach, we will demonstrate the framework by comparing 3 Pandora units in their measured slant columns of NO_2.

The comparison approach makes use of the assumption, that independently calibrated Pandoras are expected to be able to measure the same slant column amount, if other effects such as e.g., USS, bad pointing, wrong instrument handling can be excluded. Additionally, if the units would measure at exactly the same time, there should only be a difference due to a calibration error in the estimation of the slant column amount in the synthetic reference spectra. As a consequence, there should only be a systematic offset between different Pandoras, refered to as common uncertainty. Other variations than that, would refer to an independent uncertainty (random noise), and structured uncertainty which could be different for a certain pixel range.

D.2 Statistical model

D.2.1 Model formulation

The goal is to characterize daily variations as best as possible, without over-fitting the data, where the procedure is fitting a function for each day individually. In order to achieve this, we make use of a generalized additive regression model (GAM) [55, 54, 56, 28]. This kind of regression model has the advantages of a linear model framework, but is able to account for non-linear effects which depict our daily pattern of slant column variations.

When we want to compare several instruments, the model assumption is a shared smooth effect in time over the day, denoted as x, being the same for each instrument, and therefore being estimated together. However, due to expected calibration errors during the estimation of the slant column amount in the synthetic reference spectra, additional intercepts (offsets) $\beta_1, \beta_2, \beta_3$ are estimated which are instrument specific.

Consequently, we can setup a GAM, where our response variable y (slant column amount) is assumed to follow a Gaussian distribution

$$y \sim \mathcal{N}(\mu, \sigma)$$

, with σ being constant, and the expectation value μ being described by the linear predictor:

$$\mu = \beta_0 + \beta_1 + \beta_2 + s(x)$$

. Herein, $s(x)$ denotes the non-linear and smooth effect in x. The smooth effect for an individual measurement i is defined as the sum over k basis functions B_j, evaluated at the time x_i, and multiplied by its respective regression coefficient γ_j:

$$s(x_i) = \sum_{j=1}^{k} \gamma_j B_j(x_i)$$

D.2.2 Model estimation / BIC optimization

The model estimation is conducted with the R-package MGCV, which provides an ideal framework for model estimation. The optimization is based on penalized regression splines, where the default basis function are thin-plate splines since they represent the optimal smoother for any given basis dimension [54].

In principle, the MGCV package does the whole optimization in the background, including the choice of a proper smoothing term, and to use a proper number of basis functions. This choice is per default done by performing cross-validation. While the smoothing term has less of an impact, we realized that an insufficient number of basis functions is typically used for our goal, which cannot properly resolve potential real variations. On the one hand we want to follow structures that might be real, but do not want to do an over-fitting on the other side. To overcome this, we employ a search for the best number of basis functions k_{best}.

The objective criteria for choosing k_{best} is given by the Bayesian Information Criteria (BIC):

$$BIC = -2 \cdot logLik + \log(n) \cdot npar$$

This information criteria is based on the logarithmic likelihood $logLik$. Since the $logLik$ is always getting higher (better) values if more covariates or basis functions are used, the BIC incorporates the number of data points n and the number of estimated parameters $npar$ as a multiplicative penalty term. Consequently, the BIC serves as an ideal objective criteria to pick k_{best} without taking the risk of over-fitting.

D.2.3 Error quantification

Since we do not know the true value of the investigated response variable, the best way is to use a synthetic truth, referred to as 'baseline amount'. This baseline amount can be seen as the average of multiple datasets. In terms of the given GAM model, the baseline amount is defined as the smooth daily effect + the average of the individual intercepts:

$$\mu_{baseline} = s(x) + \beta_{baseline}$$

with $\beta_{baseline} = \frac{1}{m}\sum_{i=1}^{m}\beta_i$, and m the number of datasets used for the GAM model.

Note, that the MGCV package returns the intercepts $\beta_1, \cdot, \beta_m$ always to a reference instrument during the fitting process, wherefore the absolute values have to be calculated first. Reference in this sense means that the dataset of the first instrument is used as a reference.

The evaluation of the baseline allows to:

a create scatterplots in combination with a classical linear fit evaluation regarding intercept, slope and correlation coefficient.

b quantify the calibration error, which can be evaluated as the maximum difference between the intercept-difference to $\beta_{baseline}$.

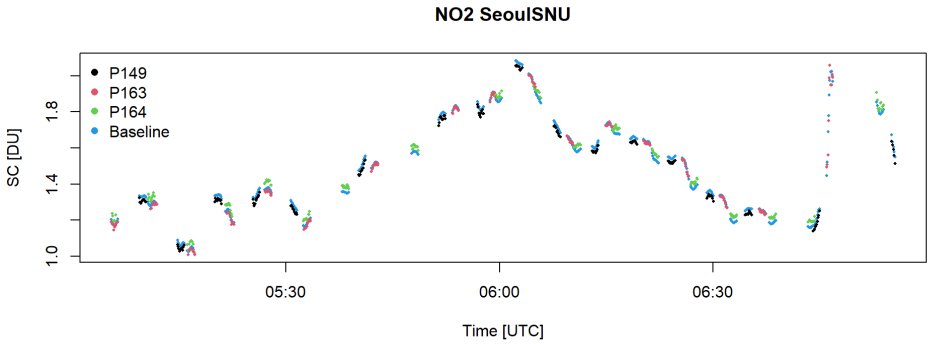
c quantify the unresolved structured uncertainty, which is given as the residual errors for the intercept-corrected values.

Note that points a-c are designed for **interComp** of slant columns amounts, since air mass factor-dependent differences are not taken into account with the current GAM formula. However, the obtained intercepts also reveals potential systematic differences, even if two datasets with air mass factor-dependences are compared.

D.3 Conceptual case study

To illustrate the concept of the GAM approach used, the following case study utilizes a rather short dataset which consists of 3 Pandora units located at Seoul, measuring direct sun total NO_2. The slant columns of the three independently calibrated units are shown in Figure 130. Just visually, there seems to be an agreement, in particular for their transition periods, while the overlapping periods indicate an offset of P164 being systematically slightly larger than P163.

Figure 130: NO_2 slant columns at Seoul for three Pandoras. Blue dots represent the estimated baseline amount.



A GAM model as specified in Section D.2 would result to a shared daily effect $s(x)$ as shown in Figure 131, if 10 or 60 basis functions are used, respectively. With $k = 10$, the model is clearly not resolving the observed features. With $k = 60$, the daily effect is already able to follow common features.

However, the objective BIC criteria reveals that $k = 110$ should be preferred, where the BIC shows an optimum (Figure 132).

Figure 131: Daily effect $s(x)$ for two different number of basis functions k used: $k = 10$ (left) and $k = 60$ (right).

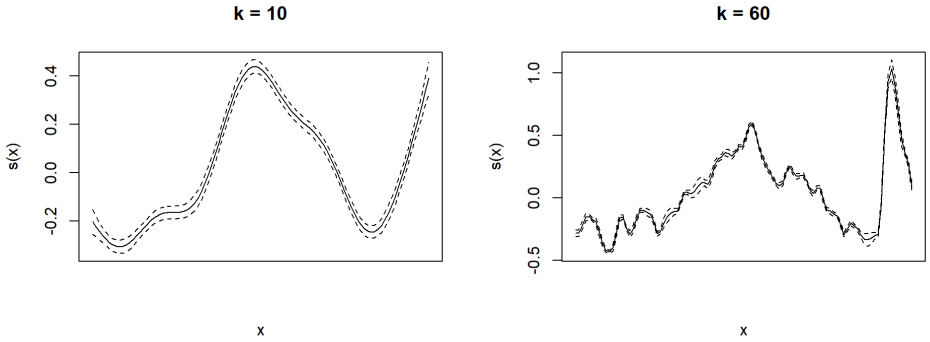


Figure 132: BIC (y-axis) as a function of the number of basis functions 'ks' (x-axis) used for the smooth daily effect. Black vertical dashed bar shows the optimum number of basis functions.

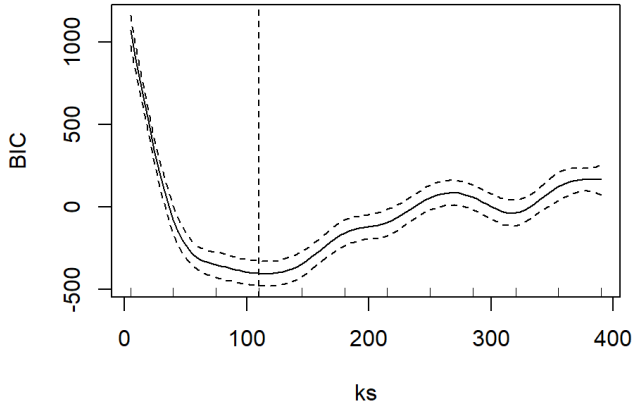


Table 16: GAM intercept difference compared to the baseline intercept, and the reported uncertainty of the MLE for NO_2 slant columns in [DU].

Intercept difference	MLE uncertainty (1-sigma)
-0.0101928	0.0211587
-0.0137408	0.0313313
0.0239335	0.0237786

The obtained intercepts for the individual instruments can be compared to $\beta_{baseline}$, and refer to a calibration error due to the field calibration with MLE (Table 16). The total range from minimum to maximum difference is approximately 0.034 DU, and can be seen as an empirical estimate about the calibration uncertainty. With the upcoming Blick version 1.8, the MLE uncertainty will be included in the calibration files, reporting the 1-sigma level calibration uncertainty. Consequently, the 2-sigma level would report already the 96% interval where the 'true' slant column value could be found. The empirical value is well within this limit, leading to the conclusion that the reported MLE uncertainty of Blick version 1.8 can be used, and that the MLE for the three Seoul units worked extremely good.

Figure 133: Scatterplots for the three instruments showing the baseline NO_2 slant column in [DU] (x-axis) against the measured value (y-axis). The legend gives the regression coefficient for a linear fit, and the R^2 value. Color-coding as in Fig.130, from left to right: P149, P163, P164

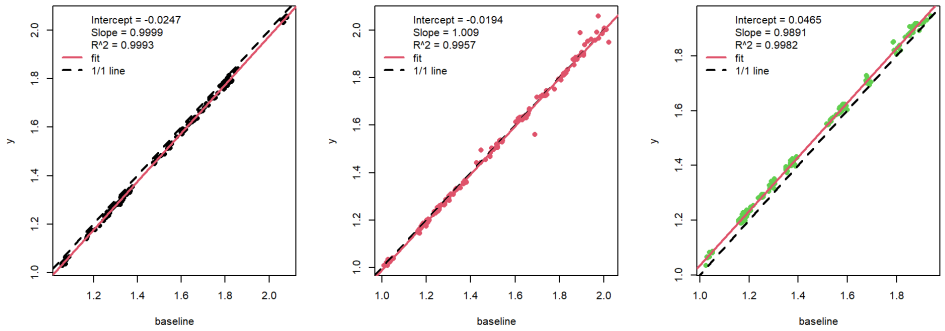


Figure 134: Overall error distribution comparing $y - \mu_{baseline}$ (left), and for the fitted GAM values, which are defined as $y - \mu$ (right) . Blue line illustrates a fitted Student's t distribution.

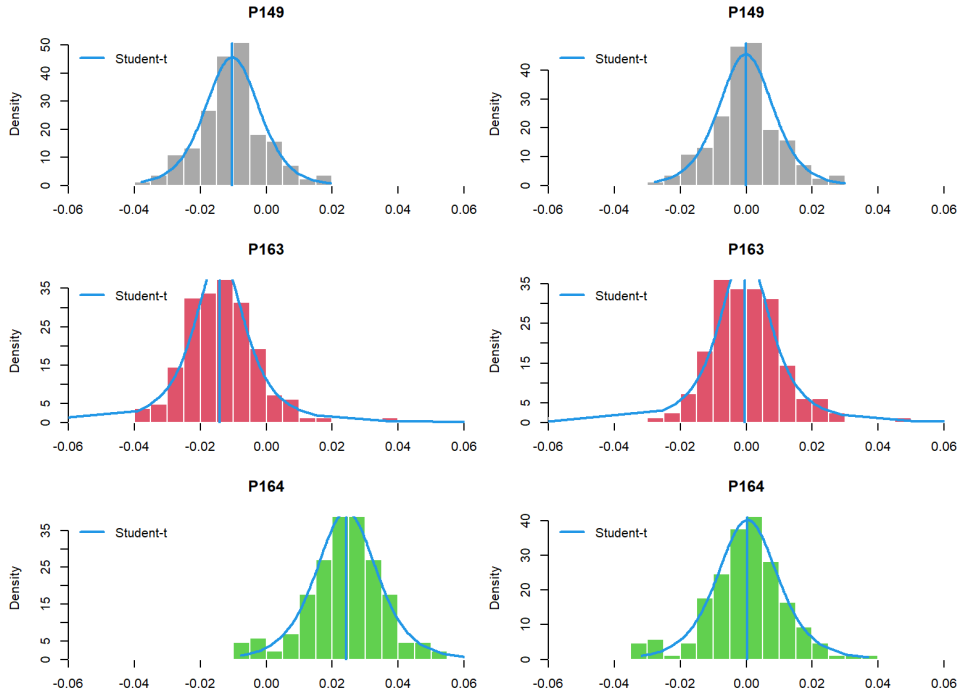


Figure 133 presents the scatter of the estimated baseline amounts versus the measured slant column amounts. The two-parameter fit, or classical linear regression, reveals that P149 is slightly lower than the baseline, and as visually already indicated, P164 is systematically larger, reported by the positive intercept. Both slope and R^2 value highlight the strong agreement with the obtained baseline amount for all three instruments. It should be noted, that this is a rather short data set, and one has to use multiple days to have a proper statistic. Therefore, differences in the slope coefficient should not be over-interpreted due to the small sample size, where individual data points can bias the intercept.

The error distribution comparing to the baseline amount, and to the fitted (intercept-corrected) values, is shown in Figure 134. In order to be able to properly account for potential outliers, the scaled Student's t distribution is added in blue [25]. The shape, and in particular the width of the error distribution, allows to quantify the structured uncertainty, which ideally is close to only the random or measured uncertainty, respectively. In this particular example, all three instruments obtain a scale parameter for the fitted Student's t distribution of 0.00873DU on average. This parameter cannot be directly compared to the 1-sigma level of the empirical standard deviation or Gaussian distribution, but the inner 68% interval width is given by a scale parameter of 0.01992DU, which translates to +/- 0.00996DU around the expected value.

D.4 Discussion and conclusions

The applied GAM model is feasible in comparing multiple datasets, in particular to account for highly non-linear daily effects. The approach is set up in a way to compare slant columns, since there is no expected air mass factor dependence. Therefore, the model introduced in D.2 is able to quantify shared daily effects if they exist. As a consequence, the model is also be able to distinguish between real variations, and random variations, based on an additional BIC optimization. Furthermore, the resulting coefficients allow a quantification of the calibration error during field calibration, and left-over structured uncertainty for the investigated data product.

It has to be mentioned, that the whole approach assumes ideal hardware conditions. An instrument which might has pointing issues, or even spectral residuals leading to a non-systematic slant column difference to the other instruments, will certainly bias the estimated baseline amount, which can further influence the final error distribution.

The introduced approach has been tested extensively for direct sun (NO_2, SO_2, HCHO, O_3) and sky data products (NO_2, HCHO), and obtains stable results regarding BIC optimization. However, there might be improvements to be made, in particular regarding the influence of biasing instruments, and if air mass factor dependent differences are expected:

- using median instead of mean regression intercepts to account for individual instruments which could bias the baseline estimation
- account for AMF-dependent differences with a new regressor
- include the information of measured (independent) uncertainty for real/random feature separation

- extending the approach towards distributional regression models, and explicitly account for location, scale, and shape parameter
- explicitly accounting for residual auto-correlation when choosing k_{best}.