LuftBlick Report 2019005

**Fiducial Reference Measurements for Air Quality**

New Algorithm & Product Development Plan

Version 4.1, 13th Jan 2021

<table>
<thead>
<tr>
<th>Name</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martin Tiefengraber</td>
<td>LuftBlick</td>
</tr>
<tr>
<td>Alexander Cede</td>
<td>LuftBlick</td>
</tr>
<tr>
<td>Manuel Gebetsberger</td>
<td>LuftBlick</td>
</tr>
</tbody>
</table>
## Contents

### Document Change Record
- 3

### Acronyms and Abbreviations
- 4

### 1 Introduction
- 4
  1.1 Applicable Documents .......................... 4
  1.2 Reference Documents .......................... 5

### 2 Processing software evolution
- 9

### 3 Processing algorithms and data products
- 9
  3.1 Development tool: SIMPLE .................... 9
  3.2 Calibration improvements ...................... 9
    3.2.1 Absolute calibration ...................... 9
    3.2.2 Matrix stray light correction ............ 9
    3.2.3 L1 wavelength displacement determination based on micro windows .................. 12
    3.2.4 L1 wavelength displacement determination based on long windows .................. 12
  3.3 Direct sun total NO2 .......................... 12
  3.4 Direct sun total O3 and O3 temperature .......... 13
    3.4.1 Direct sun total O3temp .................. 13
    3.4.2 Direct sun total O3 ....................... 18
    3.4.3 Fitting setup for O3 and O3temp .......... 23
    3.4.4 Example applications ..................... 23
    3.4.5 Operational O3 retrieval products ........ 25
  3.5 Direct sun total SO2 .......................... 26
    3.5.1 Fitting setup ............................. 27
    3.5.2 O3 cross-sensitivity ...................... 28
    3.5.3 Calibration of SO2 ....................... 28
    3.5.4 Example application ..................... 28
  3.6 Direct sun total HCHO ....................... 29
    3.6.1 Unwanted spectral signal .................. 30
    3.6.2 Fitting setup .............................. 30
    3.6.3 HCHO calibration .......................... 31
    3.6.4 Example application ..................... 31
  3.7 Direct sun retrievals of other moderately to strongly absorbing gases .......... 32
  3.8 Direct sun retrievals of other weakly absorbing gases .................. 32
  3.9 Sky data surface concentration and tropospheric NO2 columns ............ 32
    3.9.1 Surface concentration NO2 ................ 33
    3.9.2 Tropospheric NO2 columns ............... 33
  3.10 Sky data surface concentration and tropospheric HCHO columns .......... 34
  3.11 Sky data profiles ............................ 35
  3.12 Sky data retrievals of other moderately to strongly absorbing gases .......... 35
  3.13 Sky data retrievals of other weakly absorbing gases .................. 35
  3.14 Spectral AOD ................................. 35

### 4 Data product validation
- 36
  4.1 Validation strategy ............................ 36
  4.2 Direct sun O3 temperature validation .......... 36
    4.2.1 External validation ....................... 36
  4.3 Direct sun total O3 validation ................ 37
    4.3.1 Internal validation ....................... 37
    4.3.2 External validation ....................... 37
  4.4 Direct sun total SO2 validation ................ 39
    4.4.1 Internal validation ....................... 39
    4.4.2 External validation ....................... 45
  4.5 Direct sun total HCHO validation ............ 41
    4.5.1 Internal validation ....................... 41
    4.5.2 External validation ....................... 45

### 5 New data products schedule
- 47

### 6 ATBD
- 48

#### A Implications when using a literature reference for total O3 retrieval
- 49

#### B About spectral AOD from Pandora
- 53
  B.1 Introduction ................................. 53
  B.2 Basic equation ............................... 53
B.3 Description and uncertainty estimation for each parameter in equation 11 ........................................ 53
B.3.1 Air mass factors ........................................ 53
B.3.2 Molecular scattering optical depth ............... 54
B.3.3 Gas optical depth ........................................ 54
B.3.4 Signal outside the atmosphere ..................... 55
B.3.5 Measured signal ........................................ 55
B.4 Overall uncertainty estimation ........................ 57
B.4.1 Present situation ........................................ 57
B.4.2 First stage improvement ............................. 58
B.4.3 Second stage improvement .......................... 59
B.5 Conclusion .................................................. 60

C Statistical framework for dataset intercomparison 61
C.1 Introduction .................................................. 61
C.2 Statistical model ............................................ 61
C.2.1 Model formulation ....................................... 61
C.2.2 Model estimation / BIC optimization .............. 62
C.2.3 Error quantification ..................................... 62
C.3 Conceptual case study ................................. 63
C.4 Discussion and conclusions ........................... 65

Document Change Record

<table>
<thead>
<tr>
<th>Issue</th>
<th>Date</th>
<th>Section</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>30th Jun 2019</td>
<td>All</td>
<td>First version</td>
</tr>
<tr>
<td>2.0</td>
<td>14th Jan 2019</td>
<td>All, but in 3.4, 3.5, 3.6 and A</td>
<td>Major extensions in the direct sun total O3 and SO2 sections. Direct sun total HCHO section was updated too. Section calibration improvements was restructured and extended. Two appendix sections were added.</td>
</tr>
<tr>
<td>2.1</td>
<td>15th Jan 2020</td>
<td>All</td>
<td>Minor corrections and rewording.</td>
</tr>
<tr>
<td>3.0</td>
<td>28th Jun 2020</td>
<td>3.4, 3.6</td>
<td>Major extensions regarding direct sun total O3, effective O3 temperature. Direct sun total HCHO updates.</td>
</tr>
<tr>
<td>4.0</td>
<td>10th Jan 2021</td>
<td>All, but in 3.4.5, 3.5.1, 3.6.2, 4 and C</td>
<td>Section about data product validation was added. Additions to SO2 and HCHO fitting setup. New appendix section about statistical intercomparison framework.</td>
</tr>
<tr>
<td>4.1</td>
<td>13th Jan 2021</td>
<td>4.2.1</td>
<td>Added O3 temperature climatology to Figure 37</td>
</tr>
</tbody>
</table>
Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ1</td>
<td>Fitting window starting wavelength</td>
</tr>
<tr>
<td>1σLEV</td>
<td>1σ uncertainty level</td>
</tr>
<tr>
<td>2σLEV</td>
<td>2σ uncertainty level</td>
</tr>
<tr>
<td>2σUNC</td>
<td>2σ-Uncertainty</td>
</tr>
<tr>
<td>AMF</td>
<td>Air mass factor</td>
</tr>
<tr>
<td>AOD</td>
<td>Aerosol Optical Depth</td>
</tr>
<tr>
<td>ATBD</td>
<td>Algorithm Theoretical Basis Document</td>
</tr>
<tr>
<td>AXC-C</td>
<td>Auto Cross Calibration for Column</td>
</tr>
<tr>
<td>AXC-T</td>
<td>Auto Cross Calibration for Temperature</td>
</tr>
<tr>
<td>BIRA</td>
<td>The Royal Belgian Institute for Space Aeronomy</td>
</tr>
<tr>
<td>BPol</td>
<td>Background polynomial order</td>
</tr>
<tr>
<td>BSS</td>
<td>Blick Software Suite</td>
</tr>
<tr>
<td>CrossComp</td>
<td>Comparison of direct sun and sky data retrievals of a Pandora.</td>
</tr>
<tr>
<td>E-MLE</td>
<td>Extended Minimum Langley Extrapolation</td>
</tr>
<tr>
<td>ESA</td>
<td>European Space Agency</td>
</tr>
<tr>
<td>FOV</td>
<td>Field of View</td>
</tr>
<tr>
<td>FRM4AQ</td>
<td>Fiducial Reference Measurements for Air Quality</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full Width at Half maximum</td>
</tr>
<tr>
<td>H2Ov</td>
<td>Water vapor</td>
</tr>
<tr>
<td>InterComp</td>
<td>Direct sun intercomparison of co-located Pandoras</td>
</tr>
<tr>
<td>LSF</td>
<td>Line spread function</td>
</tr>
<tr>
<td>MLE</td>
<td>Minimum Langley Extrapolation</td>
</tr>
<tr>
<td>NO2</td>
<td>Nitrogen dioxide</td>
</tr>
<tr>
<td>O2</td>
<td>Oxygen</td>
</tr>
<tr>
<td>O3temp</td>
<td>Effective ozone temperature</td>
</tr>
<tr>
<td>O4</td>
<td>Oxygen dimer</td>
</tr>
<tr>
<td>OD</td>
<td>Optical Depth</td>
</tr>
<tr>
<td>Pandora</td>
<td>Pandora spectrometer system</td>
</tr>
<tr>
<td>pm</td>
<td>Pico meter</td>
</tr>
<tr>
<td>POp</td>
<td>Pandonia Operations</td>
</tr>
<tr>
<td>PSS</td>
<td>Pan Software Suite</td>
</tr>
<tr>
<td>RMS</td>
<td>Root mean square of the spectral fitting residuals</td>
</tr>
<tr>
<td>SC</td>
<td>Slant column</td>
</tr>
<tr>
<td>SIMPLE</td>
<td>SIMulate Pandora data LEvels</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise-ratio</td>
</tr>
<tr>
<td>SO2</td>
<td>Sulfur dioxide</td>
</tr>
<tr>
<td>SpatSL</td>
<td>Spatial Stray Light</td>
</tr>
<tr>
<td>SpecSL</td>
<td>Spectral Stray Light</td>
</tr>
<tr>
<td>SZA</td>
<td>Solar zenith angle</td>
</tr>
<tr>
<td>TN</td>
<td>Technical Note</td>
</tr>
<tr>
<td>USS</td>
<td>unwanted spectral signal</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
</tbody>
</table>

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 5 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 6. Note that the ATBD itself is given in Cede [6]. The section just links to it.

In the appendix section two further topics are discussed. Section A talks about implications for a O3 products using a literature reference in the retrieval and section B presents uncertainty estimations for a spectral AOD product.

1.1 Applicable Documents


1.2 Reference Documents


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 [18]) 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 therefore improve homogeneity of certain data products (as e.g. shown for total O3 in Tiefengraber and Cede [40]). 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. 49] 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. By this the full slit function can be represented by 3 sets of parameters.
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 [30] helped to reveal this issue. This is shown in figure 1, 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.

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

<table>
<thead>
<tr>
<th></th>
<th>PSS</th>
<th>BSSv1.7</th>
<th>BSSv1.8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATA LOGISTICS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Policy</td>
<td>Generic disclaimer (principal investigator not known)</td>
<td>Site application form</td>
<td>same as BSSv1.7</td>
</tr>
<tr>
<td>Processing frequency</td>
<td>Daily</td>
<td>Real time (~ 10 min.)</td>
<td>same as BSSv1.7</td>
</tr>
<tr>
<td>Station naming</td>
<td>Any name allowed</td>
<td>“Intuitive read” criteria (e.g. IBK ⇒ Innsbruck)</td>
<td>same as BSSv1.7</td>
</tr>
<tr>
<td><strong>TRACEABILITY</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calibration validity</td>
<td>Not implemented (manual renaming)</td>
<td>Validity period and version fully implemented for processing</td>
<td>same as BSSv1.7</td>
</tr>
<tr>
<td>Data files naming</td>
<td>Not uniform (dependent on calibration file)</td>
<td>Uniform description with versioning and provided/updated by PGN</td>
<td>Calibration validity period added to file name.</td>
</tr>
<tr>
<td><strong>QUALITY CONTROL, QUALITY CHECK</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw data</td>
<td>Not implemented</td>
<td>First check upon data arrival at server</td>
<td>Front-end diagnosis plots added</td>
</tr>
<tr>
<td>Retrieval products</td>
<td>Not implemented</td>
<td>Each data level categorized in high, medium, low quality data</td>
<td>Automated daily statistical break point analysis added</td>
</tr>
<tr>
<td><strong>DATA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrument characterization</td>
<td>Common characterization steps (see e.g. Cede and Tiefengraber [19])</td>
<td>major additions: matrix stray light correction, absolute calibration, micro window wavelength correction for reference spectra.</td>
<td>major additions: stray light correction includes structures (e.g. “ghost”), improved wavelength correction (up to 1 pm)</td>
</tr>
<tr>
<td>Gas reference amount calibration</td>
<td>MLE</td>
<td>MLE, E-MLE</td>
<td>MLE, E-MLE, AXC uncertainty categories added: independent, structured and common uncertainty. E.g. calibration uncertainty or impact of wrong cross section temperature considered.</td>
</tr>
<tr>
<td>Uncertainty treatment</td>
<td>propagated measurement noise</td>
<td>same as PSS</td>
<td></td>
</tr>
<tr>
<td>Climatologies, databases</td>
<td>none</td>
<td>Rayleigh AMF lookup-tables (for sky)</td>
<td>G3 profile climatology (TOMS v8 [14]), NO2 stratospheric column climatology (OSIRIS [15]), temperature/pressure climatology (COSPAR [13])</td>
</tr>
<tr>
<td>Product categories</td>
<td>Total columns from sun</td>
<td>Total columns from sun and moon, tropospheric columns from sky data</td>
<td>Total columns from sun and moon, tropospheric columns and profiles from sky data</td>
</tr>
</tbody>
</table>
Figure 1: 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.

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 2 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), and 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 2 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.

Figure 2: 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 of the certificate can serve as SpecSL parameter calibration. For comparison, the transmission calculated without SpecSL correction is shown in gray.
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 3.4.1) are considered.

3.3 Direct sun total NO2

The current operational total NO2 is already at a very high quality level. Precision-wise total NO2 performs exceptionally good (better than about 0.5 %, compare e.g. Tiefengraber and Cede [39], Zhao et al. [48]). The current way of calibrating NO2 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. [26]). 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 NO2), which further is assumed to be constant over the day. Sub-setting the data to meet these criteria is the major uncertainty source in the MLE and is directly connected to the variability of the tropospheric gas content during the calibration period.

In BSSv1.8 we have included an extended MLE technique (E-MLE) where the (variable) tropospheric content taken from sky observations (see section 3.9) 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 3. In the left figure panel the MLE analysis plot is drawn with the retrieved relative NO2 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 NO2 columns, which are provided by the BSS (see section 3.9). The MLE has also been tested at stations with different NO2 conditions (e.g. urban, suburban, remote) and has proven to work well. Therefore is has been implemented in BSSv1.8.

---

1A measured spectrum is selected as reference. For this reference spectrum the NO2 content needs to be found.

2Subset the data in air mass factor bins and define for each bin a certain percentile to be representative for the background value.
Figure 3: Total NO2 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).

3.4 Direct sun total O3 and O3 temperature

The current operational total O3 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 O3 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 O3 starting at solar zenith angle (SZA)=70 deg and sometimes even at smaller values.
- We assume a fixed effective O3 temperature (O3temp) of 225 K in the algorithm, which commonly leads to underestimation (overestimation) of total O3 in summer (winter) [48]. In addition to this bias in total O3, incorrect O3temp assumptions also bias SO2 retrievals.

Above mentioned points are discussed in more detail e.g. in Müller et al. [31] and Tiefengraber and Cede [40].

A major step to improve total O3 is the incorporation of O3temp 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 O3 cross sections is rather weak and can easily be 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 O3 and O3temp is to determine the total O3 column and O3temp in the chosen reference itself. Once this is known, the retrieval is rather straightforward.

3.4.1 Direct sun total O3temp

A bias of 1 K in O3temp translates to about 1 DU SC error and hence is significant. The annual variation of O3temp increases with latitude\(^3\) which makes seasonal biases in O3 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 O3temp to be insignificant in our columnar perspective, mainly for the reason that O3temp is strongly linked to the vertical profile of O3 which emphasizes stratospheric O3 abundances. In addition, the diurnal variation of O3 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) [38].

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 O3temp, which was not expected by us. Even a registration error in the order of some pico meters (pm) already significantly biases O3temp, as illustrated in Figure 4 (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 O3temp, 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.
Form spectrometer between 310 and 350 nm. In their study they used the Echelle spectrometer below 310 nm and the Fourier Transform (spectrometer) were used to determine the cross sections and for both a wavelength calibration accuracy of 5 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-IAS) at the 1st PGN workshop in Innsbruck, Austria [20] (and personal communications with Richard Siddans (RAL Space)) that the Serdyuchenko O3 cross sections [24] might be shifted by 3 pm\textsuperscript{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 O3temp caused by the O3 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 O3 per 1 pm error in the wavelength registration.

\textsuperscript{4}As stated in Gorshelev et al. [24], 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 O3 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 4). 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 [44], spectral features are generated. This effect is usually referred to as tilt-effect [29] 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 5 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 5:** 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. 34, 21]. 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 O3 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 6 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.

**Figure 6:** 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 O3 cross sections**

As could be seen from Figure 4 both a wavelength shift in the reference spectrum and the O3 cross sections can bias O3temps. Consequently, in order to assess a possible wavelength shift in the O3 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 O3temp exhibiting no AMF dependent diurnal variation. We hypothesize that if we would know the actual O3temp of a particular day, e.g. from O3 sondes, with all systematic retrieval errors to be sufficiently small, the only reason for a significant variation of a retrieved O3temp over the day would be a

---

\(^5\) This characterization is done by measuring the spectral dispersion as retrieved from spectral lamps for different instrument temperatures.
Following this hypothesis, 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 O3 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 7 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 O3temp 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 O3temp 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.

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 8a and 8b 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 8: Analysis of Serdyuchenko O3 cross-sections based on field data from Izana. Same explanation as for Figure 7.

(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 O3temp

Figure 9: 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 O3temp 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 O3 cross-sections. Consider a reference spectrum, chosen at a low SZA (red dot in Figure 9). This spectrum is applied to a series of test spectra at sufficiently higher AMFs (blue range in Figure 9, explanation follows later). For this setup, a series of O3temps are guessed for the reference. If for each of those reference scenarios the retrieved O3 SC for the test spectra is investigated, one can experience the SC to be biased high if the O3temp was not guessed correctly. This distinct pattern is shown for simulated data in Figure 10, where the inflection point occurs at the correct O3temp 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 O3temp 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 O3temp calibration prior to O3 calibration itself. Its worth noting that for the calibration of O3temp, one single clear sky day is sufficient.
**Figure 10:** Simulated variation of O3 SCs, retrieved from a test spectrum, for different biases in the O3temp assumed in the reference spectrum. The figure columns refer to the scenarios where also biases in the O3 SC in the reference are simulated (left -10 DU bias, middle no bias, right +10 DU bias). The minimum SCs coincide with the correct O3temp 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 O3temp 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 O3 cross-section also biases the O3temp 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 11a 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 9) and is in agreement with simulations. The finally picked O3temp for the reference should be the average over this range. Figure 11b illustrates one AXC-T pattern from this application for a SC difference of about 460 DU.

**Figure 11:** 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 O3temps for test spectra at ∆SC(test, ref) 300 to 500 (compare Figure 9).

(b) One example AXC-T pattern, taken at ∆SC(test, ref) = 462 DU. Compare Figure 11a.

### 3.4.2 Direct sun total O3

Already in Tiefengraber and Cede [40] we conceptually presented a first implementation of a novel calibration approach which takes advantage of the fact that O3 can be considered a non-linear absorber. Non-linear not by nature but due to the fact that O3 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 O3 SC amount in the reference spectrum at the correct temperature.

**Calibration approach for O3: initial AXC-C concept**

This first AXC concept as described in Tiefengraber and Cede [40], 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 O3 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 O3temp in the spectral fitting. Figure 12 shows the new retrievals from Pandora as blue line for total O3 in the top panel, O3temp in the middle panel and total SO2 in the bottom panel. Total O3 from the previous processing is given in pink. For comparison, total O3 retrievals from OMI (OMTO3) are shown in red and O3temp retrievals from the weekly ozone sondes launched at the ground station in magenta. The agreement in total O3 and O3temp between the datasets is excellent and the very low total SO2 values for Izana (what could be expected) indicate that biases due to cross correlations are not affecting the data anymore.

O3temp 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 O3 alone, but O3temp immediately reacts and is in general significantly biased. This makes the O3temp, in addition to having it as an additional atmospheric parameter, an extremely valuable calibration tracer.

An example can be seen in figure 13, where the time series from above (figure 12) 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 O3 relative to OMI is hardly recognizable, but the “jump” in O3temp (in particular after the second interaction) is striking. In addition to the bias, O3temp also shows clearly nonphysical variations over the day, which allows the detection of a calibration change even if the true O3temp is not known.
**Figure 13:** Same content as in Figure 12, but without total SO2. 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 O3, but particularly striking for O3temp.

Calibration approach for O3: new AXC-C concept
Due to the situation that the current AXC-C approach still relies on the accuracy of the Pandora total O3 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 O3) [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-Bouguer 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].

\[ \log(F_i) = \log(F_0) - A \cdot qs_i \]  
\[ \log(F_k) = \log(F_0) - A \cdot qs_k \]  
\[ \Rightarrow \log(F_i) = \log(F_k) - A \cdot (qs_i - qs_k) = \log(F_k) - A \cdot \Delta qs \]  

$\Delta qs$ in (3) is the well known differential slant column density. In case of $qs_k$ being zero (absorption free reference), $\Delta qs$ 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

\[ \log(F_i) = \log(F_0) - A \cdot (qs_i^{1+B} - qs_k^{1+B}) \approx \log(F_k) - A \cdot \Delta qs \]  

For this situation $\Delta qs$ 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 14 illustrates this bias for different guesses of O3 SC in the reference as a function of O3 SC. No impact of a wrong guess is discernible for a linear algorithm (right panel), but a non-linear algorithm “reacts” (left panel).
This effect is AMF dependent (as evident from figure 14), because the non-linearity is more pronounced for larger AMFs. However, the effect is generally rather small for weakly non-linear absorber like O3 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 O3 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 O3 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 O3 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:

$$\Delta^L qs^*_\text{syn} = qs^*_{\text{nl,syn}} - qs^*_{\text{l,syn}}$$  \hspace{0.5cm} (7)

$$\Delta^L qs^*_\text{lit} = qs^*_{\text{nl,lit}} - qs^*_{\text{l,lit}}$$  \hspace{0.5cm} (8)

$$\Delta^R \Delta^L qs^* = \Delta^L qs^*_\text{syn} - \Delta^L qs^*_\text{lit} \equiv 0$$  \hspace{0.5cm} (9)

With

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

Expression

(7) contains the “calibration sensitivity” factor $qs^*_{\text{nl,syn}}$

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

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

$\Delta^R \Delta^L qs^*$ is calculated for a number of test spectra at AMFs below $\approx 3$ (blue line in Figure 15), whereas the references is the same spectrum as for O3temp calibration (red dot in Figure 15). Again, one single clear sky day is sufficient for the O3 calibration.
This double delta approach comes with the advantage that possible biases in \( q_s^* \), 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 O3 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 O3temp needs to be assessed. First, O3temp is used to scale the O3 cross sections for the retrieval of \( q_s^* \) and hence it has to be assumed that O3temp is constant over day. This is unavoidable, because O3temp can not be retrieved using a literature reference. Second, O3temp is also used to create the synthetic reference. From simulations we see a 2 K error in the O3temp estimation to result in about 1 DU error for the O3 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 O3 absorption in the fitting.

An application of AXC-C is displayed in the Figure 16 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 \( \Delta R \Delta L q_s^* \). This is accompanied by the simulated \( \Delta R \Delta L q_s^* \), 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 \( \Delta R \Delta L q_s^* \) 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 15:** 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.

**Figure 16:** 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).
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.).

### 3.4.3 Fitting setup for O3 and O3temp

The optimal fitting setup should minimize systematic errors for O3 columns and O3temp, since both parameters are retrieved simultaneously. Hence the combined error was simulated for different fitting setups and is shown in Figure 18 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 18:** Simulated combined error for the retrieval of O3 and O3temp 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 O3 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.

### 3.4.4 Example applications

In the previous section the underlying calibration procedures (AXC-T+C) for the new O3 product was outlined. Now we want to investigate some example applications.

**Izana, Tenerife, half-year 2019**

Izana represents a perfect location for O3 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 17 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, O3 total columns from the Brewer #185 and O3temp retrievals from ozone sonde are shown in blue. The comparison is complemented by the TropOMI (reprocessed) offline O3 product, which gives beside the column values also estimations about the effective temperature [37]. We see a very good agreement to the reference datasets both for O3 columns and O3temp. The O3temp from satellite is retrieved as slant information and also uses the TOMS O3 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 O3temp.
Figure 17: Diurnal variation of total column O3 (top panel) and effective O3 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 O3 data products are displayed as red lines.

![Diurnal variation of total column O3 and effective O3 temperature](image1)

Goddard Space Flight Center, Maryland, March 2019

Figure 19: Diurnal variation of total column O3 (top panel) and effective O3 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 O3 data products are displayed as red lines.

![Diurnal variation of total column O3 and effective O3 temperature](image2)
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 17, the top panel of Figure 19 displays O3 columns and the bottom panel O3temp. Shown is Pandora 66 in blue, 67 in green and 68 in purple. Again the TropOMI (reprocessed) offline O3 data products are overlain in red.

From a first glance one can evidence a very close agreement for the O3 columns among the Pandoras. For O3temps, 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 20.

Considering uncertainties in SpecSL characterization as a main driver for O3temp biases, the observed discrepancies most likely are related to that. The impact on O3 calibration is small, due to AXC-C being sufficiently robust against smaller O3temp biases.

Figure 20: Same explanation as for Figure 19. Shown is a close up of March 12th, 2019.

In terms of satellite validation, the larger spread for O3temp does not necessarily pose a problem. Direct validation of O3temp 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 O3temp retrieval, first and foremost O3temp serves as tracer for the validity of a calibration. However, once we have a better understanding about the O3temp variation over the day, valuable implications for tropospheric O3 assessments are likely.

3.4.5 Operational O3 retrieval products

As part of the developments of the BSSv1.8 (see Table 1), the TOMS version 8 temperature and ozone profile climatology [14] was incorporated. First, this allows us a latitudinal and time dependent estimate of the effective ozone height needed for the AMF calculation. Second, the O3temp itself can be estimated as a function of time, latitude and O3 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.

The retrieval of O3temp, 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 O3/O3temp product not applicable in all cases. In these cases the O3 product based on the literature reference is a very good approximation already. In particular for low latitude stations, with rather small O3temp dynamics.

In fact, first applications turned out that the difference between the O3 product based on AXC and the (new) O3 product using the literature reference can often be explained by the difference in the O3temps (retrieved vs. climatology) only. In Figure 21 the difference in O3 columns of both products as a function of the difference between the climatological and retrieved O3temp is shown for three examples (Izana, Davos, Huelva). The column difference is seen to be within ±5 DU and clearly depends on the differences between O3temps.

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 O3temp retrieval from Pandora 121s1 in Huelva could explain this offset, but O3temp compares very well with O3sondes at Huelva. This needs further investigations.
Figure 21: Difference of Pandora O3 column retrievals between using a literature reference and O3temp climatology (LitRef) and a synthetic reference with O3temp fitting (SyntRef). This difference is shown as a function of the difference in both O3temp 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 O3 columns of both products is driven by the difference in the O3temp estimation.

Example time series (Davos, 2020) of the two new O3 products and the former product are shown in Figure 22 as diurnal mean values. The seasonal and small scale variations in O3temp, as retrieved by the AXC based O3 product are obvious (bottom panel, green dots). The impact on the O3 retrieval is striking, with bias amplitudes reaching 15 DU for the old O3 product and, as mentioned, up to 5 DU for the new O3 product based on the literature reference.

Figure 22: Comparison between the old O3 product (red) and the two new O3 products using either a synthetic reference and O3temp fitting (SyntRef, green) or a literature reference and a O3temp climatology (LitRef, blue). The top panel shows the absolute O3 column difference to SyntRef. The bottom panel shows the retrieved (SyntRef), climatological (LitRef) or fixed (old product) O3temps.

3.5 Direct sun total SO2

The strongest absorber in the spectral region of SO2 absorption is O3 and also features related to O3temp. This is why it is of importance to have O3 absorption properly corrected in a synthetic reference spectrum which is used for total SO2 retrievals.
3.5.1 Fitting setup

As suggested by sensitivity studies based on SIMPLE, the optimal fitting setup for SO2 uses a fitting window between 305 and 325 nm. The corresponding wavelength matrix plot is shown in figure 23, where relative SO2 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 SZA's, 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 23: Simulated percentage error of SO2 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 SO2) 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 SO2 fitting window, the potential additional gases would be BrO and HONO. The simulated impact of adding the gases or not is shown in Figure 24 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 SO2 errors, but this window appears to have a rather strong HONO and in particular BrO dependency at higher SZA's. 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 SZA's vanishes. If the gases would have been calibrated the error at small SZA's would still be almost zero.

Figure 24: Simulated SO2 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).
3.5.2 O3 cross-sensitivity

In the previous sections we have identified topics which could possibly limit total SO2 accuracy (O3 (+temperature) absorption, SpecSL, O3 cross section displacement). Their percentage impact was quantified using SIMPLE simulations and is summarized in figure 26. 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 O3 SC in the reference (top left), the corresponding O3temp (top right), SpecSL (bottom left) and a possible O3 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 SO2 retrieval accuracy for SZA < 70 deg. We only want to point out the importance of allowing the fitting of O3temp for total SO2 retrieval (figure panel on top right): when the O3temp was estimated incorrectly by e.g. 5 K, the impact on total SO2 is still very small as long as O3temp fitting is allowed (dark blue line). If for the same situation temperature fitting is not enabled, the emerging spectral features are wrongly explained by SO2 absorption which leads to biases (light blue line). For this case even the actual atmospheric temperature is correct in the simulation, only the O3temp 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 SO2 algorithms for ground based instruments do not take into account variable O3temp.

![Figure 26](image_url): Simulated percentage error of SO2 SCs as a function of a bias in the O3 SC in the reference (top left), O3temp in the reference (top right), SpecSL correction (bottom left) and a shift in the O3 cross sections (bottom right). For the O3temp case, also the effect is shown when no O3temp fitting is included in the fitting (faint colors).

3.5.3 Calibration of SO2

The vertical distribution and chemistry of SO2 advertise again the use of the MLE for determining the SO2 SC amount in the reference. In contrast to NO2, where a quasi constant (stratospheric) background is present, we do not expect to have a significant background for SO2.

3.5.4 Example application

Proper validation of total SO2 is not possible at the current state due to the absence of external SO2 reference datasets. Instead we make use of three co-located Pandoras at GSFC in March 2019 to compare total SO2 from those instruments (Pandora 66, 67 and 68). All three datasets have been calibrated independently and are shown in Figure 25 with errorbars. While Pandora 66 and 67 agree extremely...
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 SO2 values over the day - as one would expect. On the 19th in the morning hours, a SO2 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.

3.6 Direct sun total HCHO

Total HCHO from direct sun is currently not an operational data product 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 MAX-DOAS data, revealed this situation\(^6\). Figure 27 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 [36].

\(^6\)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 25: Diurnal variation of total SO2 at GSFC for Pandora 66 (red), 67 (blue) and 68 (green). Datasets are calibrated independently, as explained in the text.

Figure 27: 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 outgassing.
3.6.1 Unwanted spectral signal

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 [39]), 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 28, 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.

3.6.2 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 29, where relative HCHO 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. 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.

Similar as done for the SO2 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 SO2. The simulated impact of adding the gases or not is shown in Figure 30 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 additional gases would be BrO, HONO and SO2. The simulated impact of adding the gases would have been calibrated also the error at small SZAs vanishes. If the gases would have been calibrated also the error at small SZAs would increase drastically to 30 % (10 %) for BrO (HONO).

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 30:** Simulated HCHO retrieval error when either BrO, HONO or SO2 (colors) is present in the atmosphere but not fitted (left panel), or not present but fitted (right panel).

### 3.6.3 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. [35]), 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.

### 3.6.4 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 total (blue lines in Figure 32) and MAX based tropospheric (green lines in Figure 32) 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 32:** 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 from 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.
3.7 Direct sun retrievals of other moderately to strongly absorbing gases

All trace gas algorithms included in the Blick software suite use pre-convoluted cross-sections, which are given for each instrument in the Instrument Calibration File. Moderately to strongly absorbing gases like O3, O2, O4 or water vapor (H2Ov) show however either a strong slope in the cross-section across the instrument slit function (like for O3) and/or are highly structured (like H2Ov). The mathematical simplification of pre-convolution in combination with sloped and/or structured cross-sections leads to non-linearities\(^7\), which have to be taken into account in the spectral fitting process. This functionality, which is included in the BSS (see section 6.4.3 for details), is mandatory for accurate retrievals of e.g. O2 or H2Ov.

Figure 31 shows all trace gases included in the spectral fitting library of the Blick software suite given as optical depth for an assumed vertical column "standard" amount in the atmosphere as listed in the figure legend. H2Ov is displayed in light green and the large number of absorption bands are clearly visible. In principle, applying spectral fitting to all those individual bands should reveal the same H2Ov content. However, it might be that the quality of the instrument characterization varies along the spectral range and e.g. the quality of the literature cross-sections as well as the temperature sensitivity of H2Ov varies also spectrally.

\(^7\) That is a non-linear relation between the slant optical depth and the slant column amount of an absorber.

Figure 31: Optical depths of the assumed vertical standard amount of trace gases included in the spectral fitting library of the Blick software suite.

To investigate the “true” H2Ov column demands a comprehensive study which will take quite some time.

O2 and O4 are well known products in supporting aerosol estimations in sky data. For direct sun, due to the fact that O2 and O4 contents are basically only linked to the atmospheric pressure, their magnitudes are well known. This offers the opportunity to also use O2 and O4 columns as calibration tracers just like O3temp.

3.8 Direct sun retrievals of other weakly absorbing gases

We define weakly absorbing trace gases to have a peak optical depths for standard column amounts below \(10^{-3}\). Considering figure 31, quite a number of trace gases besides HCHO fall into this region. The development of retrieval algorithms for those trace gases is strongly linked to improvement in reducing the USS. The order of weakly absorbing gases which should be studied has to be decided together with ESA. During the PGN Workshop in September 2019, it was suggested that HONO should be one of the first ones to be analyzed.

3.9 Sky data surface concentration and tropospheric NO2 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 NO2 and tropospheric columns of NO2 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 tropospheric profiles for both NO2 and HCHO are planned.

3.9.1 Surface concentration NO2

The NO2 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 Lind, which are outlined in detail in Lind et. al, 2019 (in preparation). One example from the CINDI-1 instrument intercomparison campaign [32] is shown in figure 33, 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)).

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

3.9.2 Tropospheric NO2 columns

The tropospheric NO2 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 [27, 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 “Hafelekarspitze” (HAF) (compare Figure 34). 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 34: 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.

A selected time period of the comparison is shown in Figure 35 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 34). 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 35: Comparison of tropospheric NO2 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 given 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 NO2 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.

3.10 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 from sky data [41], and the surface concentration of HCHO compares well to reference data sets (Spinei, in preparation).

The before mentioned validation campaign for NO2 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 36: The difference of NO2 total columns measured by a Pandora in the valley and on the mountain top is compared to tropospheric NO2 retrieved from a sky algorithm for the valley Pandora.

3.11 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.

3.12 Sky data retrievals of other moderately to strongly absorbing gases
The challenges related to the retrieval of moderately to strongly absorbing gases are equal to those mentioned for direct sun products (compare 3.7). An additional difficulty is the spectral dependence of the AMF, which for direct sun is no issue at all. The AMF estimations are one of the biggest challenges for sky data retrievals.

3.13 Sky data retrievals of other weakly absorbing gases
Since the USS only emerges in combination with (quasi) parallel beams (direct sun), sky data retrievals are not affected by this issue. As for direct sun weakly absorbing gases, the order of importance of gases to be studied has to be decided together with ESA.

3.14 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 still have to be done in order to make spectral AOD an operational Pandora data product.
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 sun O3 temperature validation

4.2.1 External validation

Comparison datasets

Table 2: Comparison dataset(s) for total column O3temp validation.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Date</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Izana, Pandora 121s1</td>
<td>2016-12-1 to 2019-6-10</td>
<td>Carlos Torres, Alberto Redondas</td>
</tr>
</tbody>
</table>

Izana

The times series of the O3temp retrieved from Ozonesondes is shown in Figure 37 (as blue dots) and compared to diurnal mean values of O3temp 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 O3 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 O3 profiles collected at Izana, scaled to the respective sonde profiles (work thankfully done by Alberto Redondas from AEMET).

Figure 37: O3temp times series in Izana, derived from ECC Ozonesonde (blue) and Pandora (red). Pandora data are diurnal means. For comparison, O3temp derived from the TOMS v8 ozone profile climatology is also shown (in gray).

The analysis of the difference of both data sources (Pandora O3temp are diurnal means) unfolds a rather high agreement, as illustrated in Figure 38, 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 O3temp 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 37, Dec. 5th, 2018.)
4.3 Direct sun total O3 validation
4.3.1 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 39, 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 O3temps.

Figure 39: 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.

### 4.3.2 External validation

Comparison datasets

Table 3: Comparison dataset(s) for total column O3 validation.

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

In principle we would like to compare vertical total column amounts of O3. 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 40, 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 O3 layer height is set constant to 22 km for the Brewer AMF, but is adjusted according to the O3 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 O3 layer height further by the station altitude.

Figure 40: O3 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 41 summarizes the comparison results, showing the percentage difference Pandora - Brewer as a function of AMF for the two operational O3 products: The product using a synthetic reference and does O3temp fitting (left figure panels) and the product using the literature references with O3temp 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 3), 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 O3 absorption, gradually increases slant column values by up to about 1 % at AMF 5.

At smallest AMFs, potential differences due to O3 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 ±2% across the hole AMF range.
4.4 Direct sun total SO2 validation

4.4.1 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 C.

Example days of total SO2 are illustrated in Figure 42. The instruments follow common patterns, although there is also the indication that P164 is systematically lower than the other two instruments.

Figure 42: Timeseries of total column SO2 for Pandoras P149, P163, P164 at Seoul-SNU from 2020-06-13 to 2020-06-17

The GAM approach is applied on the SO2 slant columns for each day, as exemplarily shown in Figure 43. 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 43: Example days of total column SO2 for Pandoras P149, P163, P164 at Seoul-SNU from 2020-06-13 to 2020-06-17
Figure 43: 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 44), where P164 is as expected 0.0168 mmol/m² lower than the baseline, and P149 0.0155 mmol/m² larger. Therefore, the range of median intercepts is 0.0324 mmol/m², 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 44: Intercepts of the obtained daily GAM fits, relative to the baseline intercept

The scatter plot in Figure 45 reports similar intercepts, and shows $R^2$ values between 0.87-0.88.

Figure 45: Scatterplots for the three instruments showing the baseline SO2 slant column in $[\text{mmol/m}^2]$ (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.44, 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 46, and the corresponding statistics are presented in Table 4. The inner 68% interval reports an uncertainty ranging from +/- 0.020 to +/- 0.027 mmol/m².
Table 4: Summary statistics of the Student’s t distribution for the AMF errors shown in Figure 46. From left to right: Location parameter (Location), scale parameter (Scale), degree of freedom (df), inner 68% interval width (PI68), inner 95% interval width (PI95).

<table>
<thead>
<tr>
<th>Location</th>
<th>Scale</th>
<th>df</th>
<th>PI68</th>
<th>PI95</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMF [1-1.5)</td>
<td>0.00053</td>
<td>0.02042</td>
<td>21.96723</td>
<td>0.04155</td>
</tr>
<tr>
<td>AMF [1.5-3)</td>
<td>-0.00337</td>
<td>0.02638</td>
<td>16.02136</td>
<td>0.05414</td>
</tr>
</tbody>
</table>

Figure 46: 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 / m²].

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

4.5 Direct sun total HCHO validation

4.5.1 Internal validation

The InterComp is done based on the same dataset and analysis as explained in 4.4.1. Example days of total HCHO are illustrated in Figure 47.

The GAM approach is applied on the HCHO slant columns for each day, as illustrated in Figure 48. The instruments share a variable daily effect, which can be captured by the regression model, as illustrated in Figure 48 (right). An evaluation of the intercepts over all days does not reveal a significant systematic calibration error on median for the individual instruments (Figure 49), although P164 indicates to be slightly lower than the other two instruments. However, the range of median intercepts is 0.0079 mmol/m², which is an indicator that the calibration with the E-MLE approach is applicable.

Figure 47: Timeseries of total column HCHO for Pandoras P149, P163, P164 at Seoul-SNU from 2020-06-13 to 2020-06-17

Figure 48: 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 49: Intercepts of the obtained daily GAM fits, relative to the baseline intercept

![Boxplot](image)

The scatter plot in Figure 50 is supporting the intercept evaluation with its linear intercept close to zero, and additionally shows $R^2$ values between 0.92-0.96.

Figure 50: Scatterplots for the three instruments showing the baseline HCHO slant column in [mmol/m²] (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.49, from left to right: P149, P163, P164.

![Scatterplots](image)

The corresponding error distribution for all instruments together compared to the estimated baseline amount, is illustrated in Figure 51, and the corresponding statistics are presented in Table 5. The inner 68% interval reports an uncertainty ranging from +/- 0.048 to +/- 0.057 mmol/m². Since the calibration divergence is found to be 0.008 mmol/m² (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 5: Summary statistics of the Student’s t distribution for the AMF errors shown in Figure 51. From left to right: Location parameter (Location), scale parameter (Scale), degree of freedom (df), inner 68% interval width (PI68), inner 95% interval width (PI95).

<table>
<thead>
<tr>
<th>Location</th>
<th>Scale</th>
<th>df</th>
<th>PI68</th>
<th>PI95</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMF [1-2)</td>
<td>0.00059</td>
<td>0.05542</td>
<td>13.65305</td>
<td>0.11439</td>
</tr>
<tr>
<td>AMF [2-3)</td>
<td>0.00016</td>
<td>0.04500</td>
<td>7.03352</td>
<td>0.09628</td>
</tr>
<tr>
<td>AMF [3-6)</td>
<td>0.00958</td>
<td>0.05313</td>
<td>12.36208</td>
<td>0.11009</td>
</tr>
</tbody>
</table>

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 C.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 53, 54, 55 show two example days, each for MexicoCity (P142), Wakkervestroom (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 27).
Figure 51: 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/m²].

Figure 52: 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.
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 52. 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. 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. [43] 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 [22]. This might raise the question whether sky data products from ground allow a comprehensive validation of satellite derived HCHO.

### 4.5.2 External validation

#### Comparison datasets

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Date</th>
<th>PI [Publication]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mexico City, Pandora 142s1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAX-DOAS</td>
<td>2019-5-1 to 2020-4-15 (not continuous)</td>
<td>Claudia Rivera Cárdenas [16].</td>
</tr>
<tr>
<td>FTIR</td>
<td>2019-3-1 to 2019-12-1 (not continuous)</td>
<td>Claudia Rivera Cárdenas [16].</td>
</tr>
</tbody>
</table>

---

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 6) is shown in Figure 56 (MAX-DOAS in blue, FTIR in green), together with the Pandora data (in red).

**Figure 56:** 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 57, 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 57:** Same figure description as for Figure 56, 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 58 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.
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 59: 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.

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.

5 New data products schedule

The suggested schedule for new or refined data products introduced in section 3 is given in table 60. 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 60 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. Which products shall be developed, when it says "two more gases" will be
decided in the project meetings. This table will be updated with future versions of this report.

Figure 60: Suggested schedule for new data products for the first half of the project.

<table>
<thead>
<tr>
<th>Estimated release date</th>
<th>Data product(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM2</td>
<td>July 2019 Surface concentration and tropospheric NO2 from sky data</td>
</tr>
<tr>
<td>PM6</td>
<td>June 2020 Improved NO2 total columns</td>
</tr>
<tr>
<td>MTR</td>
<td>December 2020 Direct sun total O3/O3temp</td>
</tr>
<tr>
<td></td>
<td>Direct sun total SO2</td>
</tr>
<tr>
<td></td>
<td>Direct sun total HCHO</td>
</tr>
<tr>
<td></td>
<td>Surface concentration and tropospheric HCHO from sky data</td>
</tr>
<tr>
<td></td>
<td>HCHO profiles from sky data</td>
</tr>
<tr>
<td></td>
<td>NO2 profiles from sky data</td>
</tr>
<tr>
<td>PM9</td>
<td>June 2021 Direct sun total O4</td>
</tr>
<tr>
<td></td>
<td>Direct sun total NO2 from spectrometer 2</td>
</tr>
<tr>
<td></td>
<td>Spectral AOD (first stage)</td>
</tr>
<tr>
<td></td>
<td>Direct sun total HONO</td>
</tr>
<tr>
<td>PM11</td>
<td>December 2021 Direct sun total O2</td>
</tr>
<tr>
<td></td>
<td>Direct sun total H2O</td>
</tr>
<tr>
<td></td>
<td>Direct moon total N2O</td>
</tr>
<tr>
<td></td>
<td>Surface concentration and tropospheric column H2O</td>
</tr>
<tr>
<td></td>
<td>Spectral AOD (second stage)</td>
</tr>
<tr>
<td>PM13</td>
<td>June 2022 Direct sun total columns for CHOCHO</td>
</tr>
<tr>
<td></td>
<td>Profiles for the same gases</td>
</tr>
<tr>
<td>FP</td>
<td>December 2022 Direct moon total columns for NO3</td>
</tr>
<tr>
<td></td>
<td>Direct sun total columns for one more gas</td>
</tr>
<tr>
<td></td>
<td>Profiles for the same gases</td>
</tr>
</tbody>
</table>

6 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 - L1 Algorithm: 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 6.4 - L2 Fit Algorithm: 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 6.5 - L2 Algorithms: this describes the algorithms to produce the final data from the derived slant columns. Currently this includes the "Direct Algorithm" and the "Five Angles O2O2-Ratio Algorithm".
A Implications when using a literature reference for total O3 retrieval

One known source of a bias in total O3 is when a constant O3temp is assumed in the spectral fitting retrieval (see e.g. Zhao et al. [48]). At some occasions, comparisons to Brewer spectrophotometers and former comparisons to OMI [31] indicate systematic underestimation of the current standard total O3 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 O3 is shown in the top panels of Figure 61 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 61: Diurnal variation of total O3 (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 O3. Second, across the SpecSL correction scenarios, total O3 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 O3 columns appear to get. In this cases BPol seems to “explain” parts of the O3 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 O3 columns on the fitting window starting wavelengths ($\lambda_1$). The resulting O3 columns for $\lambda_1$ ranging from 300 to 310 nm...
are summarized in figure 62. The shown scenario is for BPols=2.

**Figure 62:** Diurnal variations of total O3 (top panel) for different selections of λ1 (color-coding). The evaluated BPols are shown in the bottom panel for 11:45.

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 64, 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

**Ad 1.: Dependency of BPols on total O3**

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. 33]. A study by Carlund et al. [17] 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 O3 absorption. Figure 63 shows the suggested spectral AOD from this study (red dots), compared to a regular Angstrom 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 63:** Spectral AOD as measured as campaign average in Carlund et al. [17] (red dots). Data-points 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 Angstrom parameterization is shown as dotted line.
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. [17]. Further, as expected (but not shown here), the choice of BPol has less impact on the O3 columns for the simulation using the Angstrom approximation.

Figure 64: 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 parameterization, the accordance between measurements and simulation is significantly higher.

**Ad 2.: Dependency on λ1**

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 O3 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 [42]. 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 65:** Simulated SC error for retrievals using a literature spectrum as reference. A “colored-noise” pattern (spectral feature) is added to the measurements to simulate 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 O3 SC is shown in Figure 65 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 O3 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 O3 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 66). 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 66: Same figure as 65, but the simulation was done with the new AOD parameterization as shown earlier.

Remarks when using a literature reference spectrum for total O3 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 O3 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 O3 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 O3 data.

6. The probability for that to happen starts with an RMS larger $\approx 1e^{-3}$.

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 O3 product (see section 3.4.2). By considering these remarks, the Pandora total O3 product using a literature reference ("out of the box" O3) 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 O3 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:

\[ F = F_0 \cdot \exp \left( -m_{AER} \cdot \tau_{AER} - m_{SCA} \cdot \tau_{SCA} - \sum_i m_{GASi} \cdot \tau_{GASi} \right) \]  \hspace{1cm} (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.

\[ \tau_{AER} = \frac{1}{m_{AER}} \cdot \left( \ln F_0 - \ln F - m_{SCA} \cdot \tau_{SCA} - \sum_i m_{GASi} \cdot \tau_{GASi} \right) \]  \hspace{1cm} (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
67 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. O3) and a tropospheric gas at a nominal HEFF of 4.2km (labeled TROP, e.g. HCHO). Values for m\textsubscript{AER} and m\textsubscript{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 67:** Estimated uncertainty in direct AMS as a function of SZA

![DIRECT SUN AIR MASS FACTOR UNCERTAINTY](image)

### B.3.2 Molecular scattering optical depth

The molecular scattering OD \(\tau_{SCA}\) at standard atmospheric pressure (1013.25hPa) is shown in figure 68 (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 \(\tau_{SCA}\). Pressure typically varies about 2% in time (2σLEV) around the climatological mean, which directly translates to 2σUNC≈2% for \(\tau_{SCA}\). If a pressure sensor would be available (e.g. included in the Pandora spectrometer system), the uncertainty in \(\tau_{SCA}\) would go down significantly to 2σUNC≈0.02%.

**Figure 68:** Optical depths for "typical" columns amount of trace gases in the Pandora database and for molecular scattering (MOL SCATT) as a function of wavelength.

![TYPICAL OPTICAL DEPTHS FOR 1.0nm RESOLUTION](image)

### B.3.3 Gas optical depth

Figure 68 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, \(\tau_{GAS}\) 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 69 shows the estimated $2\sigma$ UNC in the gas ODs in a linear scale, which is based on the column amount from figure 68 and the percentage retrieval $2\sigma$ UNC estimations given in the legend of figure 69. The reason for the small error in the ODs for O2O2 and O2 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 69 have no significant influence on the spectral AOD retrievals. This leaves basically O3, O2, molecular scattering and above all H2O as the gases to influence spectral AOD over the ultraviolet (UV) and visible wavelength range. Note that NO2 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 [28] has shown that even for non-pristine sites we may achieve a $2\sigma$ UNC 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% ($2\sigma$LEV). 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 70: 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 70 shows the simulated direct sun signal change due to SpecSL for different SZAs as a function of wavelength. 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-\(\alpha\)-band around 760 nm. As expected, the influence of SpecSL is smaller, when U340 is used.

While the lines in figure 70 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 70), 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 2\(\sigma\) UNC 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 71 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 71 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 \(2\sigma_{\text{UNC}}\) caused by it is half of the values in figure 71.

**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 \(2\sigma_{\text{UNC}}\) of 6%.

**Fiber bending**  In previous studies \([28]\) 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 \(2\sigma_{\text{UNC}}\) due to the fiber bending to 4% without fiber guide and 1% with fiber guide.

**Figure 71:** 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 72 and 73 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. \(F_{0\text{Cal}}\) stands for the uncertainty in the \(F_0\) calibration and \(F_0\text{Drift}\) 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 \(\tauO3\) and \(\tauH2O\) respectively.

The black line is the combined uncertainty, where all contributions are assumed to be uncorrelated, except for the pressure dependent parameters \(\tauSCA\), \(\tauO2O2\), and \(\tauO2\), which are assumed fully correlated (in the legend listed as SCA, O2, O4). REST is the combination of all error sources described before, but not singled out in the legend. The dashed black line is the “target \(2\sigma_{\text{UNC}}\)” of 0.04. This is based on the stated \(1\sigma_{\text{UNC}}\) of 0.02 for AOD by AERONET.

**Figure 72:** Estimated error contributions for spectral AOD in the present situation at SZA=0°.
As observed in figures 72 and 73, 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 $2\sigma_{UNC}=2\%$ 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 $2\sigma_{UNC}$ 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 74 for SZA 0°, 60°, 70° and 80°. It is obvious that in the present situation making an operational data product for AOD in Pandoria is not useful.

### B.4.2 First stage improvement

We believe a meaningful spectral AOD data product can be made in Pandoria 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 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 75 to 77.

**Figure 75:** Estimated error contributions for spectral AOD after the first stage improvement at SZA=0°.

**Figure 76:** Estimated error contributions for spectral AOD after the first stage improvement at SZA=70°.

**Figure 77:** Estimated error in spectral AOD after the first stage improvement at SZA=0°, 60°, 70° and 80°.

As seen in figure 77, 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,O2,O4 by a factor of 100.
- A water vapor retrieval is developed with a $2\sigma$ UNC of 6% for tH2O.

The effect of these improvements is shown in figures 78 to 80. 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 78: Estimated error contributions for spectral AOD after the second stage improvement at SZA=0°.

Figure 79: Estimated error contributions for spectral AOD after the second stage improvement at SZA=70°.

Figure 80: 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 $2\sigma$ UNC 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 Statistical framework for dataset intercomparison

C.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 O3. To illustrate the approach, we will demonstrate the framework by comparing 3 Pandora units in their measured slant columns of NO2.

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.

C.2 Statistical model

C.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) [46, 45, 47, 25]. 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 $\sigma$ being constant, and the expectation value $\mu$ 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 $\gamma_j$:

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

### C.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 [45].

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.

### C.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 $\beta_1, \beta_2, \beta_m$ 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.
C.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 NO2. The slant columns of the three independently calibrated units are shown in Figure 81. 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 81: NO2 slant columns at Seoul for three Pandoras. Blue dots represent the estimated baseline amount.

A GAM model as specified in Section C.2 would result to a shared daily effect \( s(x) \) as shown in Figure 82, 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 83).

Figure 82: Daily effect \( s(x) \) for two different number of basis functions \( k \) used: \( k = 10 \) (left) and \( k = 60 \) (right).

Figure 83: 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 7: GAM intercept difference compared to the baseline intercept, and the reported uncertainty of the MLE for NO2 slant columns in [DU].

<table>
<thead>
<tr>
<th>Intercept difference</th>
<th>MLE uncertainty (1-sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.0101928</td>
<td>0.0211587</td>
</tr>
<tr>
<td>-0.0137408</td>
<td>0.0313313</td>
</tr>
<tr>
<td>0.0239335</td>
<td>0.0237786</td>
</tr>
</tbody>
</table>

The obtained intercepts for the individual instruments can be compared to $\beta_{\text{baseline}}$, and refer to a calibration error due to the field calibration with MLE (Table 7). 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 84: Scatterplots for the three instruments showing the baseline NO2 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.81, from left to right: P149, P163, P164

Figure 85: Overall error distribution comparing $y - \mu_{\text{baseline}}$ (left), and for the fitted GAM values, which are defined as $y - \mu$ (right). Blue line illustrates a fitted Student’s t distribution.

Figure 84 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 85. In order to be able to properly account for potential outliers, the scaled Student’s t distribution is added in blue [23]. 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.

C.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 C.2 is able to quantify shared daily effects if they exist. As a consequence, the model is also 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 (NO2, SO2, HCHO, O3) and sky data products (NO2, 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 $k_{best}$. 