CEOS Intercalibration of Ground-Based Spectrometers and Lidars

Minispectrometer Intercalibration and Satellite Validation

Report 7

Validation report

	Name	Function	Date
prepared by	Martin Tiefengraber	LuftBlick	2014-02-20
	Alexander Cede	LuftBlick, CSO	2014-02-20
checked by	Katherine Cede	LuftBlick	2014-02-20
approved by			

Contents

Document Change Record **2**

1 Introduction **3**
1.1 Reference Documents . . . 3
1.2 Definitions, Acronyms and Abbreviations . . . 4

2 Validation Data and Locations **5**
2.1 Effective Validation Location . . . 6

3 Ozone Validation **6**
3.1 Data Filtering . . . 6
3.1.1 Filter Parameter Sensitivity Test . . . 6
3.1.2 Filter Thresholds Selection . . . 9
3.2 Validation Results . . . 10
3.2.1 Validation Procedure . . . 10
3.2.2 Time Series . . . 11
3.2.3 Individual Locations . . . 13
3.2.4 Conclusions of Ozone Validation . . . 23

4 Nitrogen Dioxide Validation **25**
4.1 Conclusions of Nitrogen Dioxide Validaton . . . 25

Document Change Record

Issue	Date	Page	Observations
1	2014-01-23	All	First draft version
2	2014-01-25	All	Editing
3	2014-01-27	All	Minor changes
4	2014-02-20	All	Minor changes, paragraphs added in response to ESA comments

1 Introduction

This report is deliverable D7 of project [RD01, RD02]. Section 2 introduces the data sets considered in this validation study. Section 3 presents the comparison results for O_3 vertical columns derived from both SCIAMACHY and Pandora data, including a sensitivity test on data filtering. Section 4 focuses on the comparison of SCIAMACHY and Pandora NO_2 vertical column retrievals.

1.1 Reference Documents

No	Description
RD01	Inter-calibration of ground-based spectrometers and Lidars – Minispectrometer Intercalibration and Satellite Validation [Statement of Work], Issue 1, Revision 0, GMES-CLVL-EOPG-SW-13-0001, 15 January 2013.
RD02	Inter-calibration of ground-based spectrometers and Lidars – Minispectrometer Intercalibration and Satellite Validation [Proposal], Contract: 22202/09/I-EC, RFQ/3-12340/08/I-EC, 22 January 2013.
RD03	Inter-calibration of ground-based spectrometers and Lidars – Minispectrometer Intercalibration and Satellite Validation, Report 1: List of minispectrometers considered in this activity, 10 April 2013.
RD04	Inter-calibration of ground-based spectrometers and Lidars – Minispectrometer Intercalibration and Satellite Validation, Report 2: Recommendations for Inter-Calibration of minispectrometer networks, 25 September 2013.
RD05	Inter-calibration of ground-based spectrometers and Lidars – Minispectrometer Intercalibration and Satellite Validation, Report 4: Minispectrometer Data Quality Report, 29 November 2013.
RD06	Inter-calibration of ground-based spectrometers and Lidars – Minispectrometer Intercalibration and Satellite Validation, Report 6: Validation protocol, 29 November 2013.
RD07	Tzortziou M., J.R. Herman, A. Cede, C.P. Loughner, N. Abuhassan, and S. Naik, Spatial and temporal variability of ozone and nitrogen dioxide over a major urban estuarine ecosystem, J. Atmos. Chem., DOI 10.1007/s10874-013-9255-8, 2013.
RD08	Reed, A.J., et al., Effects of local meteorology and aerosols on ozone and nitrogen dioxide retrievals from OMI and pandora spectrometers in Maryland, USA during DISCOVER-AQ 2011, J Atmos Chem, DOI 10.1007/s10874-013-9254-9, 2013.
RD09	Herman, J., A. Cede, E. Spinei, G. Mount, M. Tzortziou, and N. Abuhassan, NO2 column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, J. Geophys. Res., 114 (D13307), 10.1029/2009JD011848, 2009.

1.2 Definitions, Acronyms and Abbreviations

No	Description
AU	Autumn
CC	Correlation Coefficient
CF	Cloud Fraction
DQ0	Pandora High Quality Data
DU	Dobson Units
ESA	European Space Agency
ESRL	Earth System Research Laboratory
FMI	Finnish Meteorological Institute
FSU	Frostburg State University
GEFS	Global Ensemble Forecasting System
GSFC	Goddard Space Flight Center
IZO	Izana Atmoshperic Research Center
NO_2	Nitrogen Dioxide
O_3	Ozone
MEDD	Median Difference
NASAHQ	National Aeronautics and Space Administration Head Quarter
NOAA	National Oceanic & Atmospheric Administration
OMI	Ozone Monitoring Instrument
PSD	Physical Science Division
SCIAMACHY	SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY
SERC	Smithsonian Environmental Research Center
SP	Spring
std	Standard Deviation
SU	Summer
SZA	Solar Zenith Angle
TO3	Ozone Total Vertical Column
TNO2	Nitrogen Dioxide Total Vertical Column
UMBC	University Of Maryland Baltimore County
UMCP	University Of Maryland College Park
USNA	Naval Academy
WI	Winter

2 Validation Data and Locations

The goal of this study is to validate the vertical column amounts of O_3 (TO3) and NO_2 (TNO2) retrieved from SCIAMACHY with the recalibrated Pandora database following the strategy outlined in RD06. Pandora characteristics and data quality are described in detail in RD04, RD05, RD06. As outlined in RD01, SCIAMACHY overpass data for each location are provided by ESA, extracted from Level 2 data by Alessandro Burini. Details on SCIAMACHY and Pandora data sets are given in Table 1.

Data set	Version number	Time period	Data format
SCIAMACHY	SCIA-OL/5.02	2009-05-03 to 2012-04-07	GEOMS H5
Pandora	001	2009-05-03 to 2013-03-31	GEOMS H5

Table 1: Details on SCIAMACHY and Pandora data sets.

Figure 1 is an overview of validation locations given in RD03. Locations symbolized by green dots have overlap with the SCIAMACHY time series, those with red dots do not. In some instances there were various Pandora units distributed within only a few tens of kilometers or even side by side. If Pandora units are deployed side by side, their data have been averaged for the validation study, which reduces both noise and potential systematic errors. If Pandora units are located apart from each other, but are still within a SCIAMACHY footprint, their values are averaged too. In some cases both situations (side by side AND within a footprint) occur. To illustrate this we introduce a classification with symbols α, β and γ (Table 2). We will attach the symbol to each validation location. If no symbol is attached, it means there is only one Pandora in the SCIAMACHY footprint. Note that situations β and γ depend on the SCIAMACHY footprint for each overpass. E.g. on some days there might be three Pandoras inside the footprint, on other days only two.

Symbol	Abbreviation
NONE	Only one single Pandora is used for the validation.
α	Multiple Pandora units are deployed at the same location at the same time.
β	Multiple Pandora units are located apart from each other but still within the range of a SCIAMACHY footprint at the same time.
γ	α and β

Table 2: Different classifications for a validation location.

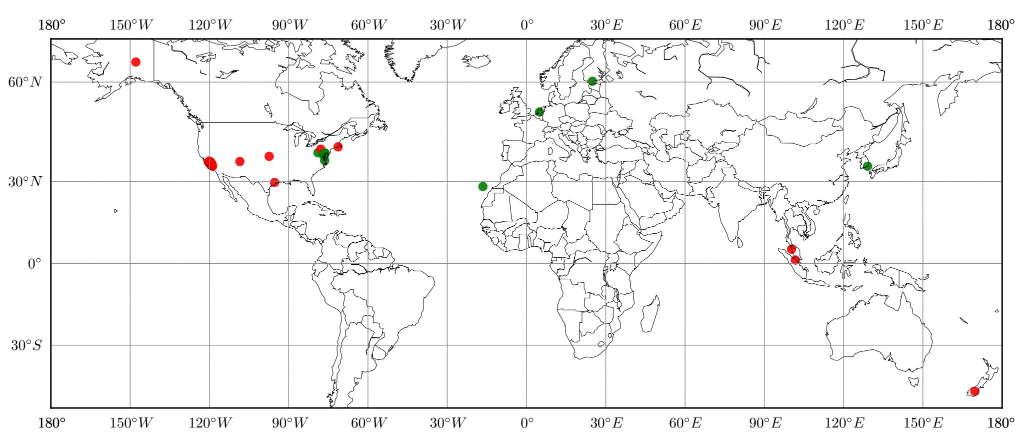


Figure 1: Validation locations. Green dots indicate locations with temporal overlap with SCIAMACHY. Locations marked in red have no overlap.

Two validation locations have a particular importance in this study. Langley (Fairfax, Virginia) as a location considered to be rural and GSFC$_\gamma$ (Goddard Space Flight Center, Greenbelt, Maryland) as an urban location. Langley provides the longest Pandora time series. GSFC$_\gamma$, so to say the Pandora headquarter, offers the densest time series with multiple Pandoras at the site in many occasions.

2.1 Effective Validation Location

The overpass time of SCIAMACHY at about 10:00 local time implies that the direct-sun data from Pandora has been measured at moderate to high solar zenith angles (SZA). If we estimate the effective O_3 (NO_2) absorption layer height to 20.1 km (1.0 km), the effective validation location is not the ground geolocation itself. Instead, the effective location is shifted towards the South-East direction (on the Northern Hemisphere). While this has a minor effect for TNO2 validation, the spatial shift can be significant for TO3 validation: in winter months in the Northern hemisphere the latitudinal (longitudinal) shift is about 40 km (15 km) for mid latitudes. In order to enhance the representativeness of the Pandora direct-sun retrievals we refer to the effective Pandora locations throughout this validation study.

3 Ozone Validation

3.1 Data Filtering

3.1.1 Filter Parameter Sensitivity Test

To estimate the influence of different filter parameters (maximum SCIAMACHY cloud fraction, Pandora vertical column uncertainty, etc.) on the TO3 comparison we performed a sensitivity study. We started with a set of thresholds for each filter parameter based on literature [RD05, RD07, RD08]. Table 3 summarizes these initial thresholds (values $<$ the threshold are accepted).

Only Pandora retrievals considered as DQ0 (high quality data) were included. The sensitivity study was only conducted for the parameters marked by an asterisk in Table 3, since the effects of all other parameters are included in them. We only varied one filter parameter threshold at the time, leaving all other parameter thresholds fixed. For each variation, we extracted TO3 from SCIAMACHY and Pandora and evaluated correlation coefficient (CC) and median difference [MEDD, median($TO3_{SCIAMACHY}$-$TO3_{Pandora}$)].

Filter parameter	Abbreviation	Threshold
Geolocation filter		
lon. footprint center distance to Pandora *	dLON	0.4° [≈ 39.2 km (LAT 28°) to ≈ 22.1 km (LAT 60°)]
lat. footprint center distance to Pandora *	dLAT	0.2° (≈ 22.2 km)
footprint area *	SciaARE	$(1.0°)^2$ (≈ 2.5×10^4 km^2)
Quality filter		
SCIAMACHY		
cloud fraction *	SciaCF	0.2
vertical column uncertainty	SciaUQ	5 DU
Pandora (DQ0)		
vertical column uncertainty *	PanUQ	5 DU
normalized rms of spectral fitting residual	PanWRMS	2.0×10^{-2}
wavelength shift	PanWVL	2.0×10^{-1} nm
direct sun air mass factor	PanAMF	5
enhanced data scatter	PanSCAT	1.0×10^{-2}
Pandora data averaging time		
data averaging time around SCIAMACHY OVP *	PanAVG	120 min

Table 3: Initial filter parameter thresholds for the sensitivity study. Pandora thresholds correspond to the DQ0 criteria. The dLON threshold range in km is based on latitudes for locations with temporal overlap with SCIAMACHY (green dots in Figure 1).

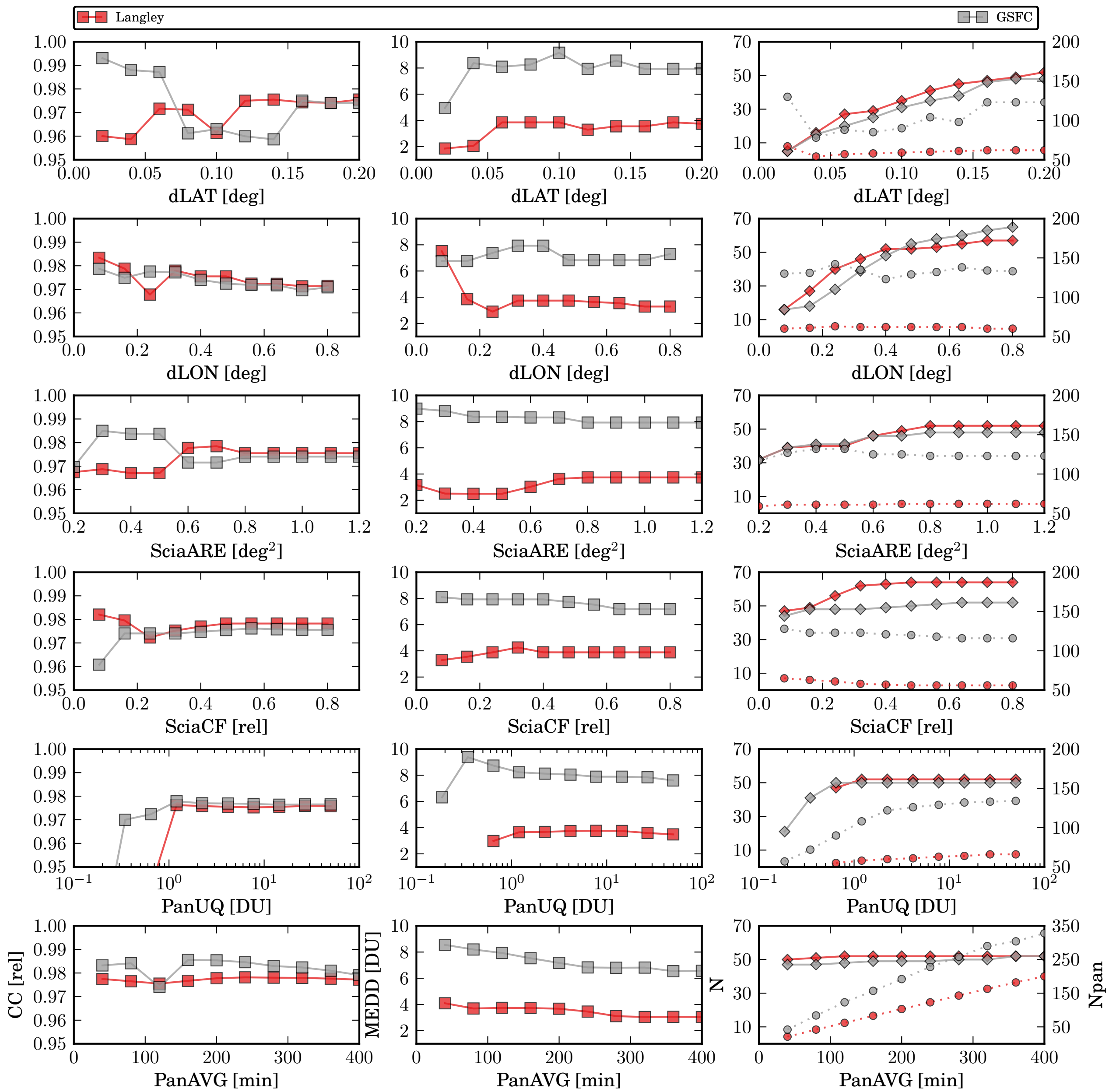


Figure 2: TO3 sensitivity study for SCIAMACHY and Pandora comparison based on data from Langley (rural) and GSFC$_\gamma$ (urban). Filter parameters are introduced in Table 3. In the left (center) panels the CC (MEDD) are plotted as a function of the threshold for each filter setup. The number of measurement points N (left axis, diamonds) and the mean Pandora data points averaged around the SCIAMACHY measurement Npan (right axis, circles) are shown in the right panels.

The results of the sensitivity test are presented in Figure 2, where for each filter setup the CC (left column) and MEDD (center column) are depicted. The panels on the right show the number of comparison data N (diamonds) and the mean number of Pandora data points averaged around each SCIAMACHY measurement Npan (circles). Npan is larger at GSFC$_\gamma$ than at Langley, since there were always multiple Pandoras deployed.. We performed the sensitivity analysis based on the data from Langley and GSFC$_\gamma$. First, those locations exhibit the longest time series leading to strong statistics. Second, these locations represent both a rural environment (Langley) and a urban environment (GSFC$_\gamma$).

Figure 2 shows the influence of chosen filter thresholds is very small for both MEDD and CC. Excluding the most stringent filter setups the MEDD (CC) variation is almost always less than 1 DU (0.03) throughout the test intervals.

The geolocation filter parameters (dLAT, dLON, SciaARE) are especially dependent on the surroundings of the location. Most likely due to the SCIAMACHY footprint geometry, dLAT affects CC stronger than dLON. In addition, dLAT does not show a clear dependency on the chosen threshold, while dLON indicates a small decrease in CC with increasing distance. The CC based on a varying footprint area (SciaARE) also reveals no clear pattern, but it seems that allowing also larger footprints flattens out the CC variation. A very stringent cloud fraction filtering for SCIAMACHY data (SciaCF) does not necessarily enhance the correlation to the Pandora direct-sun data. Although a small increase of CC is shown for the rural location (Langley), the validation for the urban environment ($GSFC_\gamma$) seems to benefit allowing higher cloud fractions. One reason for the seemingly small influence of SciaCF is that Pandora cloud filtering (PanUQ) already omits footprints with high cloud fraction. Said in other words: if the ground data reveal a sun-free situation for 120 min around SCIAMACHY overpass time, the SciaCF is automatically very small. CC actually does not depend on PanUQ if the most stringent thresholds are omitted (truncated in the Figure). The CC as a function of PanAVG is flat for the rural location (Langley) and shows a maximum between 160 and 200 min for the urban location ($GSFC_\gamma$).

3.1.2 Filter Thresholds Selection

Due to the confined temporal overlap of SCIAMACHY and Pandora data, we need to balance between stringent filtering on the one side and stronger statistics on the other. Since the overall effect of the filter parameter and thresholds on MEDD is negligible, the decision is based on CC only.

Since no clear influence of the geolocation filter parameters is seen, we allow all SCIAMACHY data, where the footprint includes the ground location in order to get the best statistics.

Regarding the cloud fraction filtering for SCIAMACHY we chose a compromise between data purity and statistics. Allowing 50 % cloud fraction only slightly reduces CC but makes more data available for the comparison. The major factor to ensure clear sky conditions is PanUQ. The choice of the PanUQ threshold within the given test interval is not critical for two reasons. First, the immense quantity of the Pandora retrievals provides enough data to maintain the number of possible SCIAMACHY data. Second, CC remains almost constant if the PanUQ threshold exceeds approximately 1 DU.

Equal to PanUQ threshold, PanAVG does not alter the number of data points to validate and hence has no influence on the strength of the statistics. Based on CC the best averaging time for Pandora data would between 160 and 200 min for urban and rural locations, respectively.

The final filter thresholds applied for the entire validation analysis are summarized in Table 4. Parameters marked by an asterisk have been optimized by the sensitivity test.

Sensitivity parameter	**Abbreviation**	**Threshold**
Geographical filter		
lon. footprint center distance to Pandora *	dLON	entire footprint
lat. footprint center distance to Pandora *	dLAT	entire footprint
footprint area *	SciaARE	-
Quality filter		
SCIAMACHY		
cloud fraction *	SciaCF	0.5
vertical column uncertainty	SciaUQ	5 DU
Pandora (DQ0)		
vertical column uncertainty *	PanUQ	5 DU
normalized rms of spectral fitting residual	PanWRMS	2.0×10^{-2}
wavelength shift	PanWVL	2.0×10^{-1} nm
direct sun air mass factor	PanAMF	5
enhanced data scatter	PanSCAT	1.0×10^{-2}
Pandora data averaging time		
data averaging time *	PanAVG	200 min

Table 4: Optimized filter parameter thresholds applied for the TO3 validation. Thresholds marked by an asterisk are the conclusion from the sensitivity test.

3.2 Validation Results

3.2.1 Validation Procedure

The TO3 validation has been carried out as follows:

- SCIAMACHY validation data preparation

  1. For each ground location, the SCIAMACHY overpass file provided by ESA was reduced to those entries, where the Pandora geolocation in included in the satellite footprint.
     → SV0 data.
  2. Data quality filtering according to Table 4 is applied to the SV0.
     → SV1 data.

- Pandora validation data preparation

  3. For each footprint in SV1 spatially and temporally overlapping Pandora units are identified (compare Table 2) and combined.
     → PV0 data.
  4. Data quality filtering according to Table 4 is applied to PV0.
     → PV1 data.

- Validation procedure

  5. PV1 data points are averaged around SV1 footprint measurement times. The averaging time is PanAVG from Table 4.
     → PV2 data.
  6. SV1 is validated against PV2.

3.2.2 Time Series

Currently, the standard Pandora TO3 algorithm does not include a variable effective O_3 temperature, while satellite products from SCIAMACHY and OMI do. As a consequence we see an undulating difference between the satellite and Pandora data during the year. Overestimation in O_3 effective temperature leads to overestimation in TO3 and vice versa. This behavior was already shown in report RD04, where a comparison of Pandora data to OMI was performed to test Pandora's calibration stability. The time series for the difference in SCIAMACHY and Pandora for Langley and $GSFC_\gamma$ again reveals the expected temperature wave, more obvious for OMI (Figure 3 bottom) than for SCIAMACHY (Figure 3 top) due to the higher data density.

Table 5 lists the seasonal median differences in TO3 (MEDD) for SCIAMACHY (second column) and for OMI (third column). The temperature related overestimation of Pandora TO3 in winter and underestimation in summer is clearly shown, as expected. The spring and autumn TO3 median differences are almost equal. Pandoras gas temperature related error is approximately 8 DU for SCIAMACHY and about 6.5 DU for OMI.

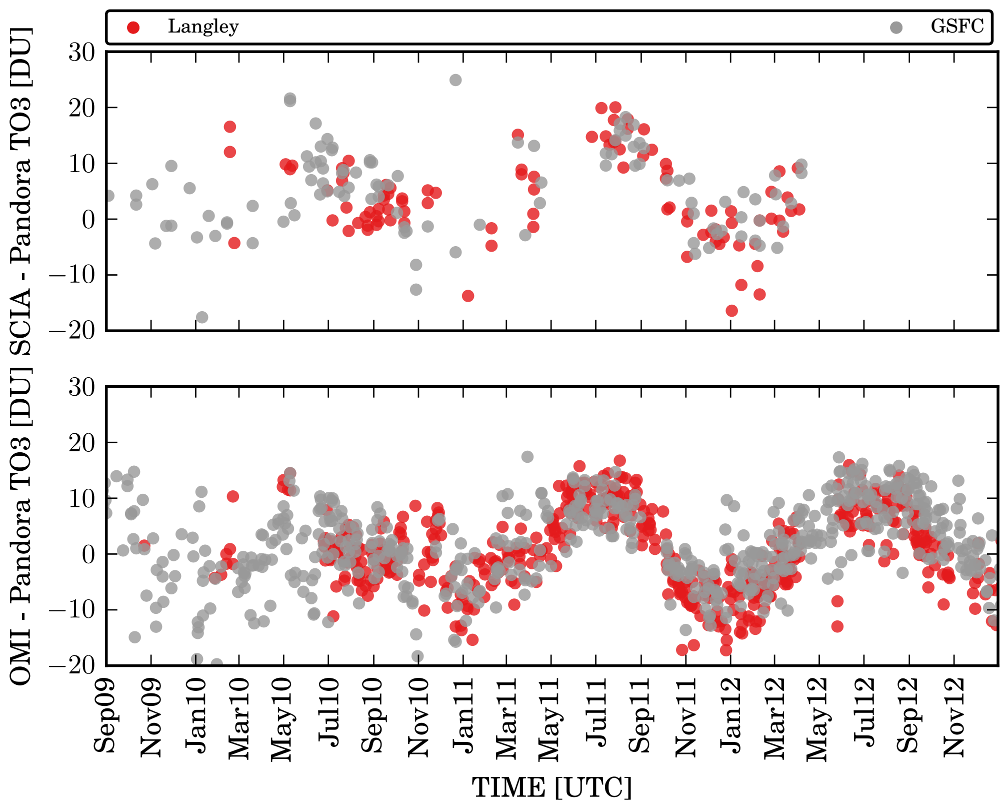


Figure 3: Time series for the difference of SCIAMACHY (top panel) and OMI (bottom panel) to Pandora. Depicted are data from Langley and $GSFC_\gamma$. The wave structure is related to the effective O_3 temperature. Due to the 6-days overpass repeat cycle of SCIAMACHY it provides significantly less coincidences than OMI.

Season	SCIAMACHY MEDD	OMI MEDD
Winter	-2.34	-4.35
Sprint	4.88	2.10
Summer	14.14	9.00
Autumn	4.91	-0.16

Table 5: Seasonal MEDD for SCIAMACHY (second column) and OMI (third column) based on Langley and $GSFC_\gamma$ data.

We can observe in figure 3 that the seasonal cycle in the TO3 differences is not as defined before May 2011 than after this date. We carefully checked whether this could be due to the Pandoras measuring at $GSFC_\gamma$ and Langley. We have no indication that they might not have worked properly. Therefore we suspect that this may be caused by the effective O_3 temperature itself. The effective O_3 temperature T_{eff} is defined in equation (1), where N is the number of vertical layers, T_i the mean temperature in layer i and $n_{O_3,i}$ the partial ozone column in layer i.

$$T_{eff} = \frac{\sum_{i=1}^{N} T_i \cdot n_{O_3,i}}{\sum_{i=1}^{N} n_{O_3,i}} \tag{1}$$

We made a rough estimation of T_{eff} at $GSFC_\gamma$ for the period 2010 to 2012 using the O_3 profile from the US Standard Atmosphere in combination with the NOAA ESRL/PSD GEFS forecast version 2, which gives the atmospheric temperature profile for 11 pressure layers (data available from `http://www.esrl.noaa.gov/psd/forecasts/reforecast2/download.html`).

Figure 4 shows the TO3 differences OMI to Pandoras (top panel) and T_{eff} (bottom panel) at $GSFC_\gamma$. The year 2012 exhibits a pronounced annual cycle in the difference OMI to Pandora, corresponding to a clear variation in T_{eff}. The offset is mainly driven by the algorithm temperature error from Pandora data. In contrast, the year 2010 shows a dampened amplitude in T_{eff} attributable to warmer winter periods. This causes on the one hand less temperature error from Pandora retrievals but on the other hand enhanced temperature errors from OMI retrievals due to enhanced deviation from the assumed temperature climatology. This superposition of different temperature related errors from Pandora and OMI leads to increased data scattering in the TO3 difference. This is clearly mirrored from November 2010 to May 2011. The OMI to Pandora offset is particularly uncorrelated in this region and coincides with rapidly varying T_{eff} around the same temperature.

Further, the outliers for example around January 2012 concur with high T_{eff}. Usually Pandora overestimates T_{eff} in winter months, however, during this period the higher actual T_{eff} minimizes the difference to the assumed retrieval temperature and hence the error. Thus this outliers are related to underestimation in the actual T_{eff} in the OMI retrieval.

A correction for the temperature related retrieval error in Pandora data is necessary to actually quantify the temperature related algorithm error when using a temperature climatology in satellite retrievals.

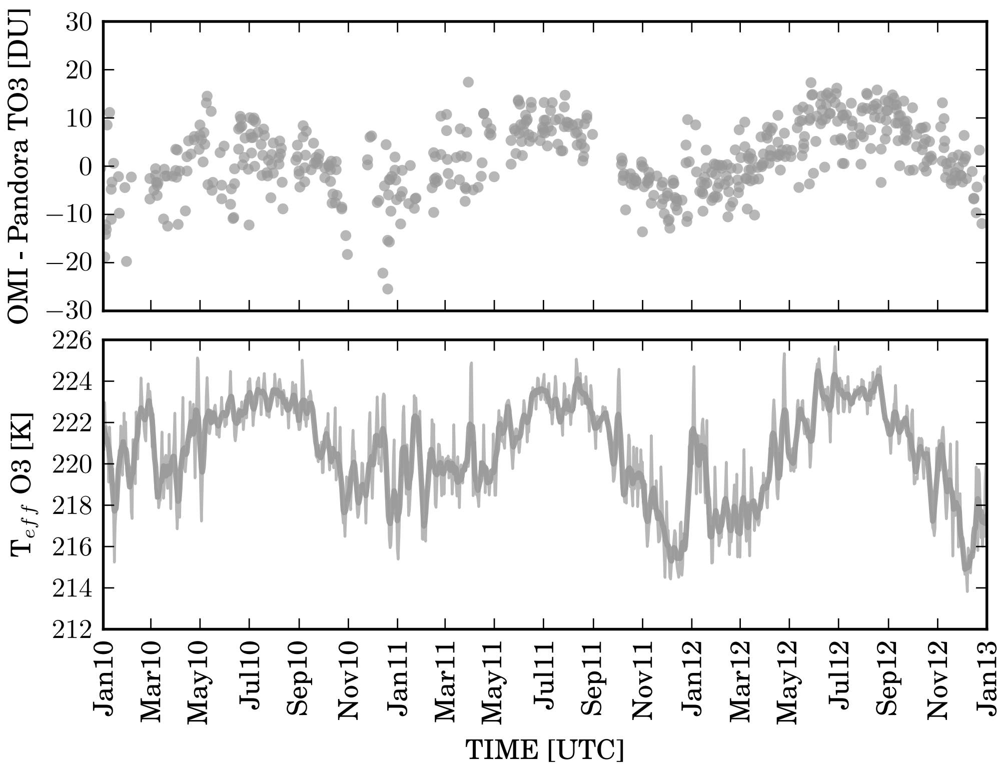


Figure 4: Depicted are the TO3 differences OMI to Pandora at $GSFC_{\gamma}$ (top panel, same as bottom panel in Figure 3) and the effective O_3 temperature T_{eff} (bottom panel, thin line daily values and thick line weekly averages). T_{eff} is based on equation (1) and makes use of temperature profiles from ESRL/PSD GEFS reforcast database and integrated profiles of O_3 number density.

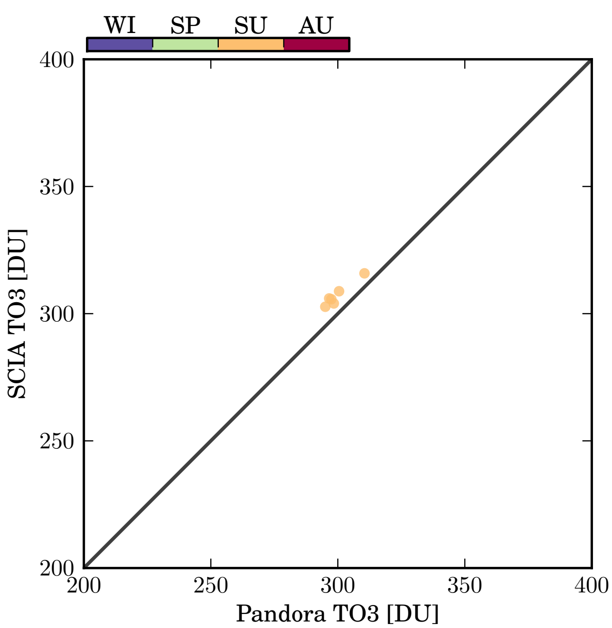
3.2.3 Individual Locations

In this section we present the SCIAMACHY to Pandora validation for TO3 for each Pandora location have time-overlap with SCIAMACHY. The data comparison is depicted as a scatter plot, where the color code indicates the season (WI = winter, SP = spring, SU = summer, AU = autumn). On the right hand side of each plot additional information is given as explained in Table 6.

Short Description	Long Description
longitude, latitude	ground geolocation of the Pandora unit (not its effective location)
n, nWI, nSP, nSU, nAU	number of data points, winter only, spring only, summer only, autumn only
max Pan.	maximum number of Pandora units involved
max Pan.loc. / footp.	maximum number of Pandora locations overlapping with the SCIAMACHY footprint
CC	correlation coefficient SCIAMACHY, Pandora
slope	slope term in the linear fit and its uncertainty (Pandora and SCIAMACHY uncertainties are included in the fit)
offset	offset term in the linear fit and its uncertainty (Pandora and SCIAMACHY uncertainties are included in the fit)
std	standard deviation of the difference SCIAMACHY and Pandora
mean	arithmetic mean of the difference SCIAMACHY and Pandora
median	median value for SCIAMACHY and Pandora difference
25-75 percentile range	25 to 75 percentile range for SCIAMACHY and Pandora difference
10-90 percentile range	10 to 90 percentile range for SCIAMACHY and Pandora difference
min-max range	minimum value to maximum value range for SCIAMACHY and Pandora difference
SCIAMACHY CF	percentage of SCIAMACHY data passing the SciaCF filter
Pandora DQ0	percentage of Pandora data passing the DQ0 filter

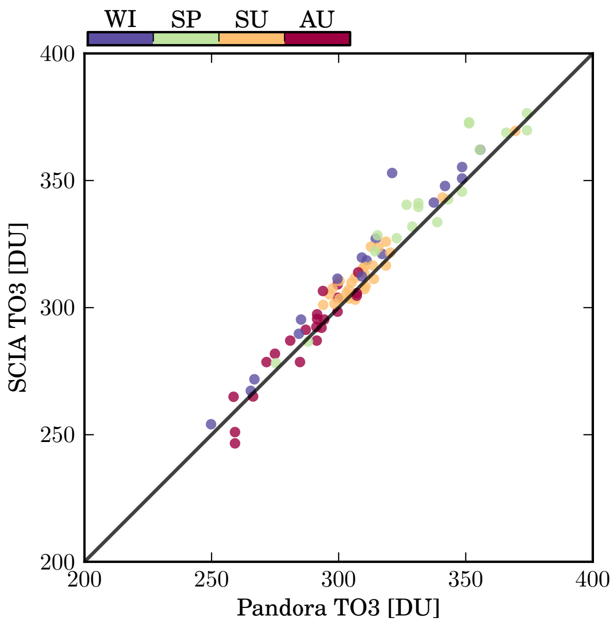
Table 6: Description on the additional information given for each location to validate.

Aldino$_\gamma$



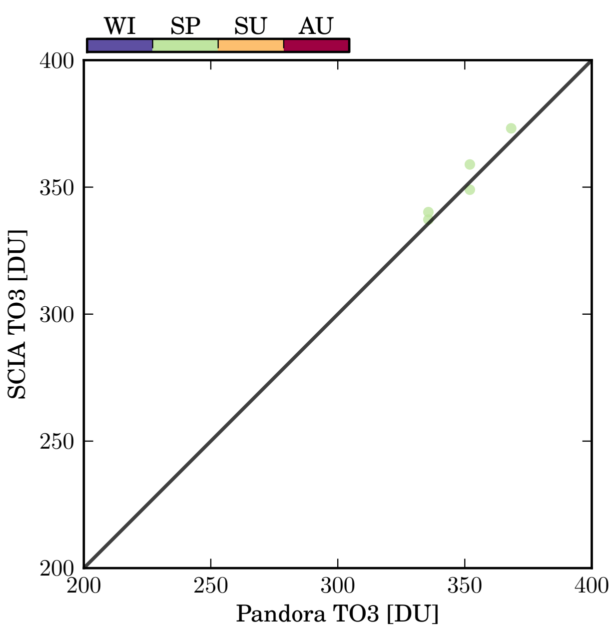
longitude, latitude	-76.203889, 39.563333
n, nWI, nSP, nSU, nAU	6, 0, 0, 6, 0
max Pan.	3
max Pan.loc. / footp.	2
CC	0.96
slope	1.23 ± 0.01
offset	-84.14 DU ± 2.33 DU
std	1.47 DU
mean	7.43 DU
median	14.94 DU
25-75 percentile range	13.11 DU to 15.31 DU
10-90 percentile range	12.48 DU to 15.87 DU
min-max range	12.37 DU to 16.40 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	80.18 %

Beltsville$_\gamma$



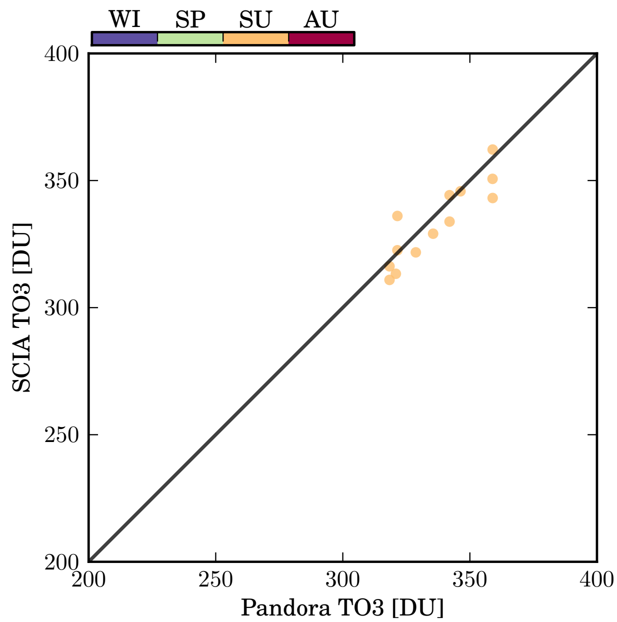
longitude, latitude	-76.878333, 39.055277
n, nWI, nSP, nSU, nAU	85, 19, 18, 26, 22
max Pan.	20
max Pan.loc. / footp.	6
CC	0.97
slope	1.02 ± 0.00
offset	-9.87 DU ± 0.48 DU
std	7.69 DU
mean	4.22 DU
median	4.45 DU
25-75 percentile range	-1.22 DU to 10.07 DU
10-90 percentile range	-4.12 DU to 14.26 DU
min-max range	-17.59 DU to 24.92 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	56.44 %

Busan



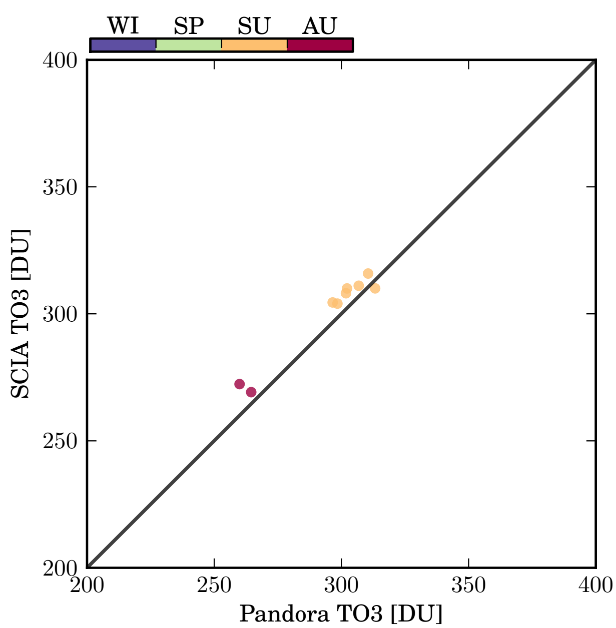
longitude, latitude	129.0825, 35.2353
n, nWI, nSP, nSU, nAU	5, 0, 5, 0, 0
max Pan.	1
max Pan.loc. / footp.	1
CC	0.97
slope	0.97 ± 0.01
offset	6.42 DU ± 1.96 DU
std	3.45 DU
mean	2.98 DU
median	4.51 DU
25-75 percentile range	1.55 DU to 4.88 DU
10-90 percentile range	-1.18 DU to 6.11 DU
min-max range	-3.00 DU to 6.94 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	67.83 %

Cabauw$_{\alpha}$



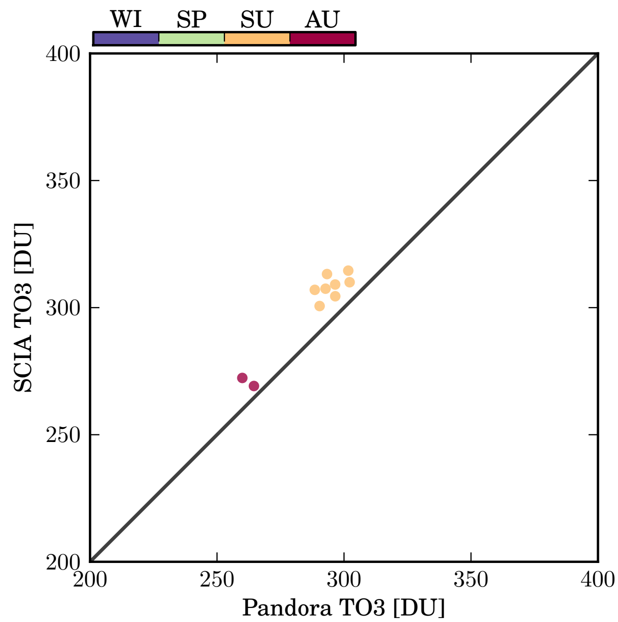
longitude, latitude	4.92625, 51.9704
n, nWI, nSP, nSU, nAU	13, 0, 0, 13, 0
max Pan.	2
max Pan.loc. / footp.	1
CC	0.89
slope	1.16 ± 0.00
offset	-57.44 DU ± 1.43 DU
std	7.26 DU
mean	-3.25 DU
median	0.55 DU
25-75 percentile range	-0.57 DU to 8.09 DU
10-90 percentile range	-1.24 DU to 10.03 DU
min-max range	-8.84 DU to 21.58 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	32.75 %

Edgewood$_{\gamma}$



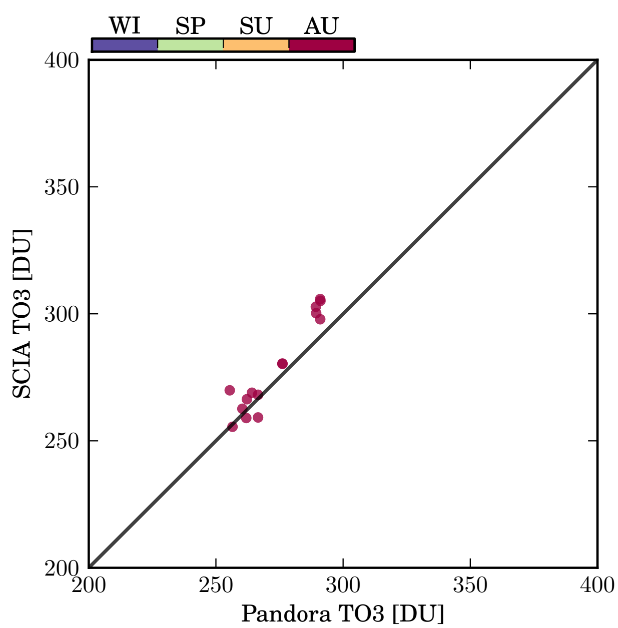
longitude, latitude	-76.29698, 39.41008
n, nWI, nSP, nSU, nAU	9, 0, 0, 7, 2
max Pan.	7
max Pan.loc. / footp.	5
CC	0.97
slope	0.96 ± 0.00
offset	2.48 DU ± 1.45 DU
std	3.86 DU
mean	5.66 DU
median	12.37 DU
25-75 percentile range	11.24 DU to 13.31 DU
10-90 percentile range	4.45 DU to 14.81 DU
min-max range	3.77 DU to 14.92 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	74.36 %

Essex$_\gamma$



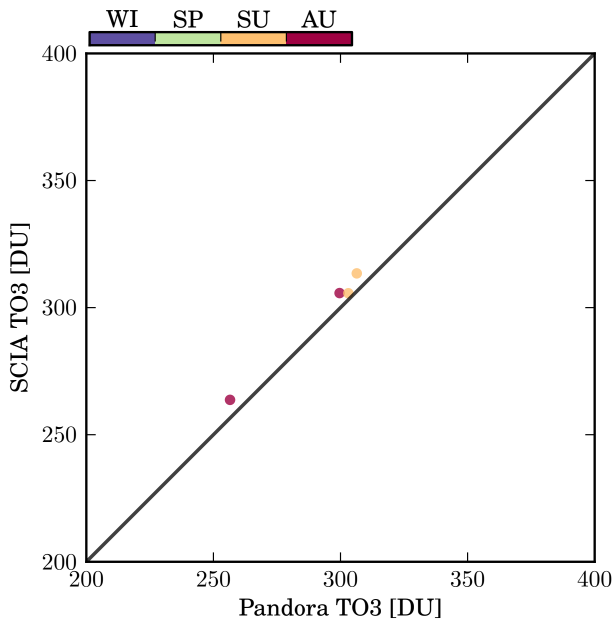
longitude, latitude	-76.474444, 39.310833
n, nWI, nSP, nSU, nAU	10, 0, 0, 8, 2
max Pan.	6
max Pan.loc. / footp.	5
CC	0.94
slope	0.75 ± 0.00
offset	58.68 DU ± 1.09 DU
std	6.21 DU
mean	12.15 DU
median	18.40 DU
25-75 percentile range	14.81 DU to 21.25 DU
10-90 percentile range	11.60 DU to 25.64 DU
min-max range	4.63 DU to 26.89 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	72.78 %

FMI



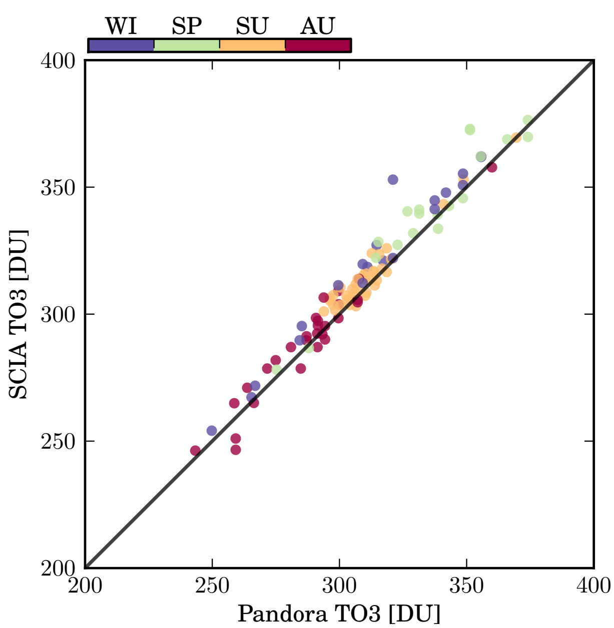
longitude, latitude	24.9612, 60.2037
n, nWI, nSP, nSU, nAU	18, 0, 3, 0, 15
max Pan.	1
max Pan.loc. / footp.	1
CC	0.99
slope	1.02 ± 0.00
offset	-10.24 DU ± 0.99 DU
std	9.11 DU
mean	5.01 DU
median	4.23 DU
25-75 percentile range	-0.36 DU to 12.84 DU
10-90 percentile range	-7.78 DU to 14.56 DU
min-max range	-10.63 DU to 25.24 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	46.61 %

FSU$_\gamma$



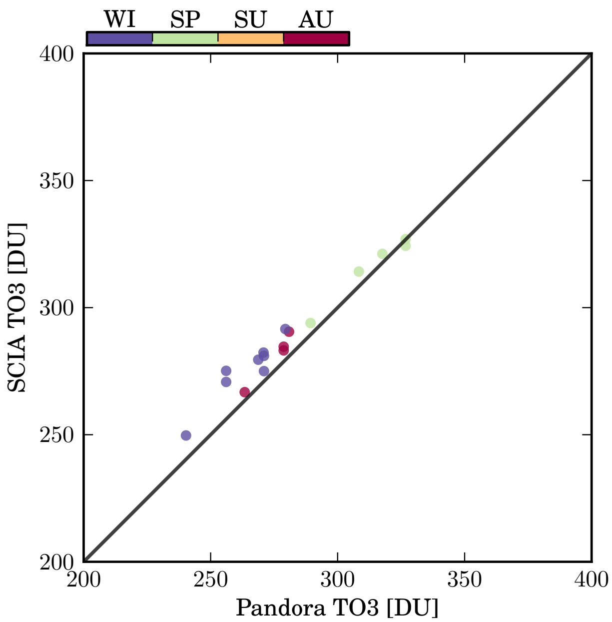
longitude, latitude	-78.9329, 39.6511
n, nWI, nSP, nSU, nAU	4, 0, 0, 2, 2
max Pan.	19
max Pan.loc. / footp.	7
CC	0.99
slope	0.95 ± 0.01
offset	5.45 DU ± 2.00 DU
std	3.07 DU
mean	5.72 DU
median	8.37 DU
25-75 percentile range	6.90 DU to 10.70 DU
10-90 percentile range	6.41 DU to 12.72 DU
min-max range	6.08 DU to 14.07 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	30.81 %

GSFC$_\gamma$



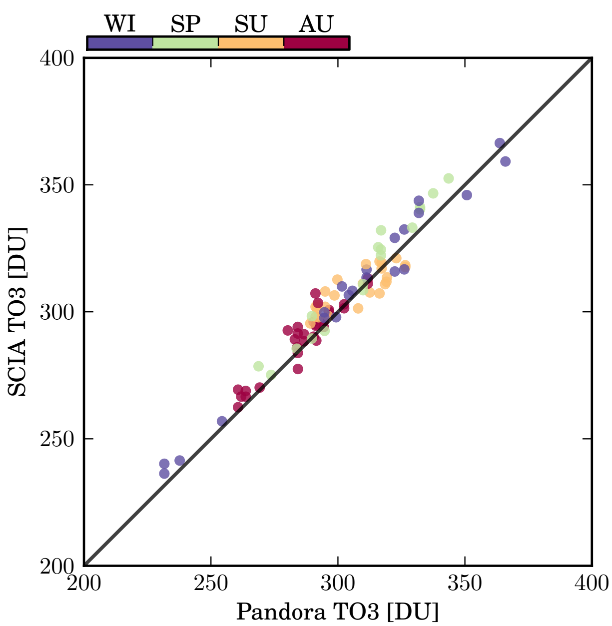
longitude, latitude	-76.8396, 38.9926
n, nWI, nSP, nSU, nAU	106, 22, 19, 37, 28
max Pan.	20
max Pan.loc. / footp.	6
CC	0.97
slope	1.03 ± 0.00
offset	-13.09 DU ± 0.43 DU
std	7.43 DU
mean	4.05 DU
median	4.94 DU
25-75 percentile range	-0.96 DU to 10.09 DU
10-90 percentile range	-4.07 DU to 13.93 DU
min-max range	-17.59 DU to 24.92 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	56.44 %

IZO



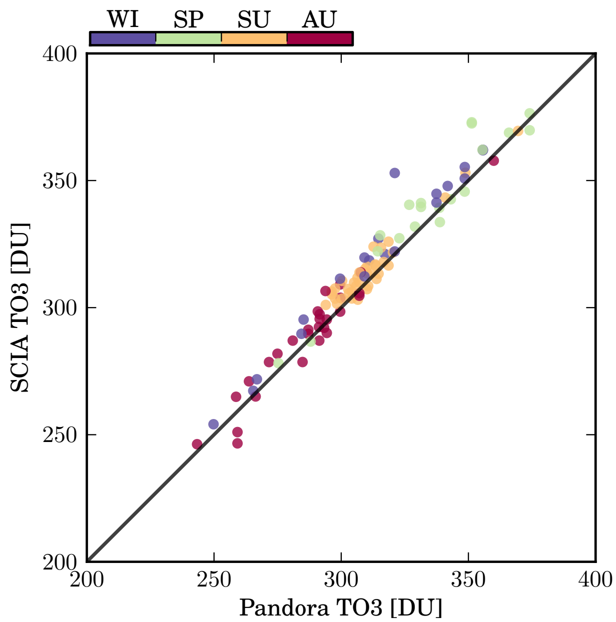
longitude, latitude	-16.4994, 28.309
n, nWI, nSP, nSU, nAU	17, 8, 5, 0, 4
max Pan.	1
max Pan.loc. / footp.	1
CC	0.99
slope	1.05 ± 0.00
offset	-19.03 DU ± 1.11 DU
std	3.66 DU
mean	7.44 DU
median	4.40 DU
25-75 percentile range	3.03 DU to 5.83 DU
10-90 percentile range	-0.87 DU to 8.46 DU
min-max range	-2.99 DU to 12.03 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	79.00 %

Langley



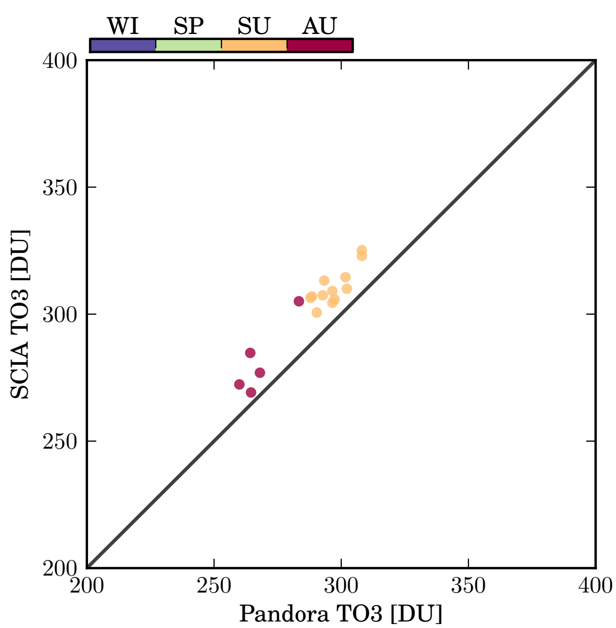
longitude, latitude	-76.3868, 37.1036
n, nWI, nSP, nSU, nAU	95, 23, 17, 25, 30
max Pan.	1
max Pan.loc. / footp.	1
CC	0.97
slope	1.05 ± 0.00
offset	-18.04 DU ± 0.47 DU
std	7.48 DU
mean	3.64 DU
median	2.09 DU
25-75 percentile range	-0.69 DU to 9.03 DU
10-90 percentile range	-4.38 DU to 14.45 DU
min-max range	-16.41 DU to 20.03 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	52.86 %

NASAHQ$_\gamma$



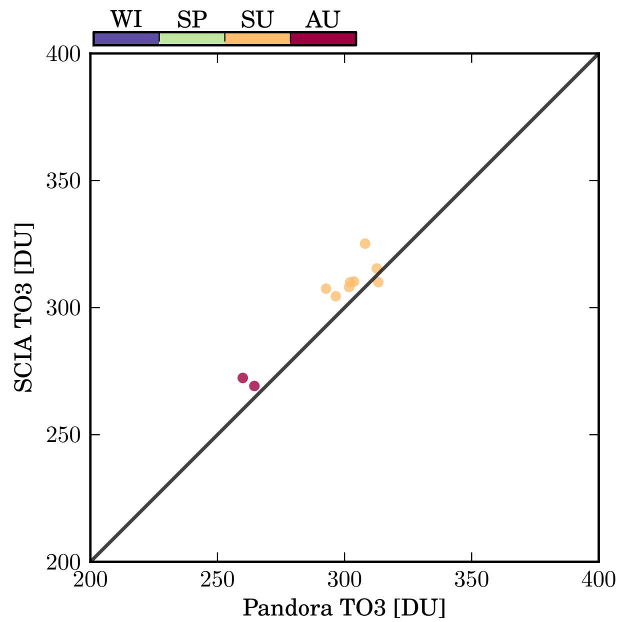
longitude, latitude	-76.8396, 38.9926
n, nWI, nSP, nSU, nAU	106, 22, 19, 37, 28
max Pan.	20
max Pan.loc. / footp.	6
CC	0.97
slope	1.03 ± 0.00
offset	-13.09 DU ± 0.43 DU
std	7.43 DU
mean	4.05 DU
median	4.94 DU
25-75 percentile range	-0.96 DU to 10.09 DU
10-90 percentile range	-4.07 DU to 13.93 DU
min-max range	-17.59 DU to 24.92 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	56.44 %

Oldtown$_\gamma$



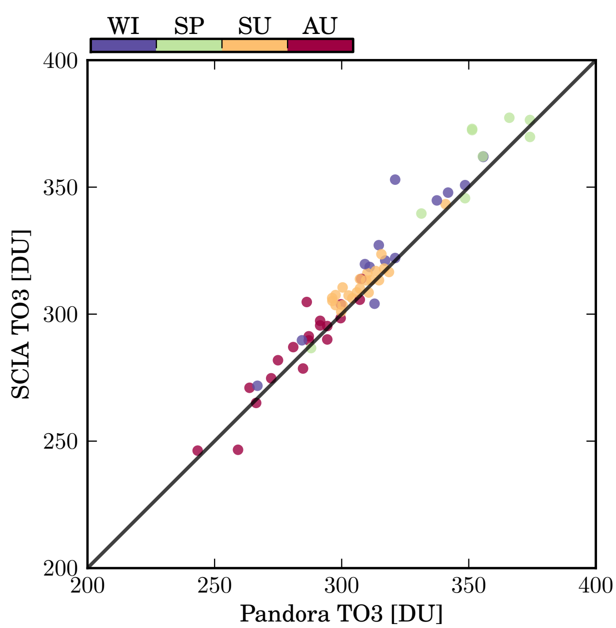
longitude, latitude	-76.596456, 39.291492
n, nWI, nSP, nSU, nAU	17, 0, 0, 12, 5
max Pan.	7
max Pan.loc. / footp.	5
CC	0.95
slope	0.78 ± 0.00
offset	47.42 DU ± 0.89 DU
std	5.93 DU
mean	13.61 DU
median	19.87 DU
25-75 percentile range	14.92 DU to 21.80 DU
10-90 percentile range	10.98 DU to 25.48 DU
min-max range	4.63 DU to 26.89 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	52.96 %

Padonia$_\gamma$



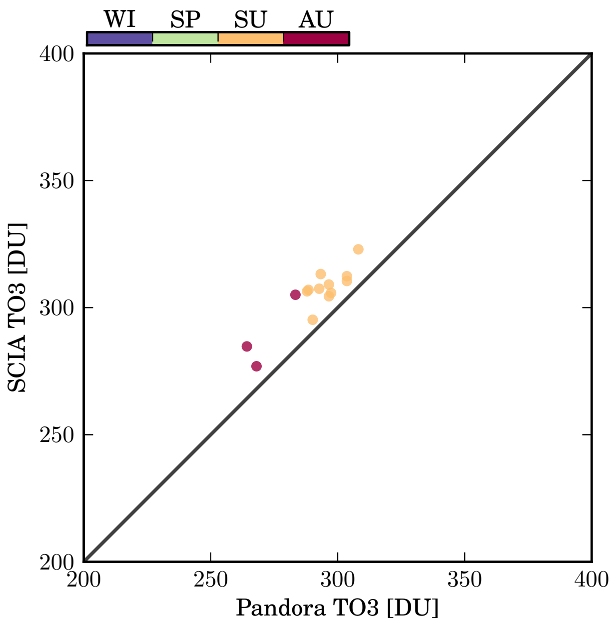
longitude, latitude	-76.631111, 39.460833
n, nWI, nSP, nSU, nAU	10, 0, 0, 8, 2
max Pan.	6
max Pan.loc. / footp.	5
CC	0.94
slope	0.92 ± 0.00
offset	10.68 DU ± 1.33 DU
std	6.10 DU
mean	7.69 DU
median	13.43 DU
25-75 percentile range	10.39 DU to 14.88 DU
10-90 percentile range	4.54 DU to 21.95 DU
min-max range	3.77 DU to 24.10 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	74.36 %

SERC$_\gamma$



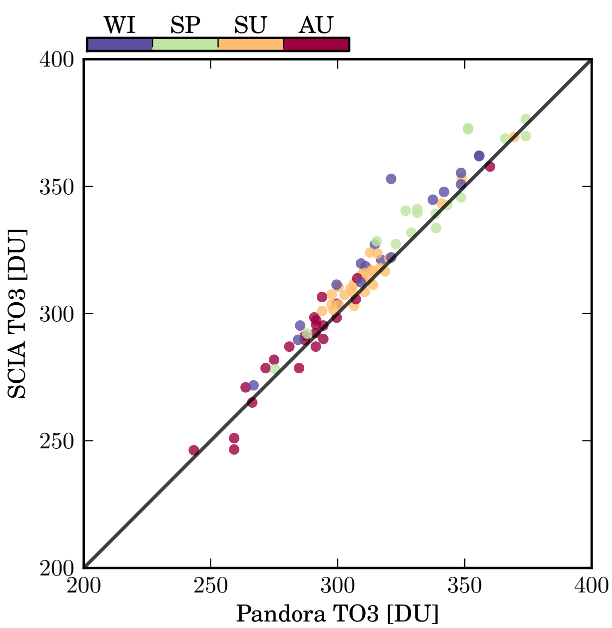
longitude, latitude	-76.55, 38.88
n, nWI, nSP, nSU, nAU	65, 15, 9, 22, 19
max Pan.	16
max Pan.loc. / footp.	6
CC	0.97
slope	1.01 ± 0.00
offset	-6.89 DU ± 0.54 DU
std	8.01 DU
mean	4.59 DU
median	5.54 DU
25-75 percentile range	-1.01 DU to 10.17 DU
10-90 percentile range	-3.89 DU to 16.50 DU
min-max range	-15.82 DU to 24.92 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	56.44 %

UMBC$_\gamma$



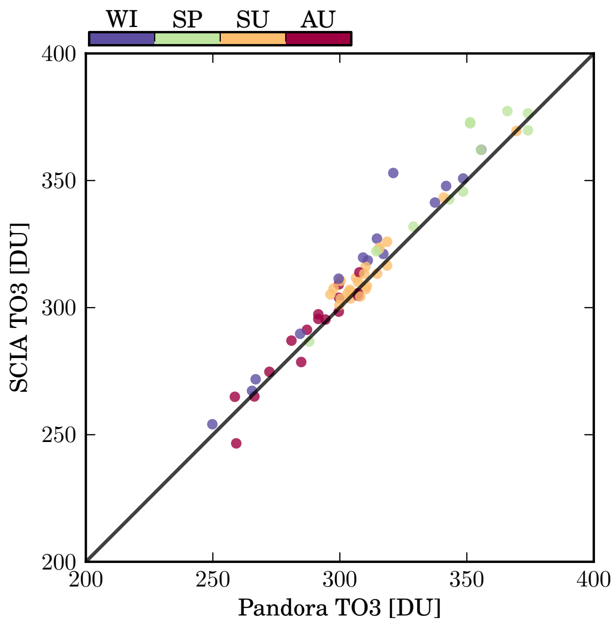
longitude, latitude	-76.709157, 39.2547
n, nWI, nSP, nSU, nAU	14, 0, 0, 11, 3
max Pan.	7
max Pan.loc. / footp.	5
CC	0.89
slope	0.96 ± 0.00
offset	-7.85 DU ± 1.23 DU
std	5.26 DU
mean	13.36 DU
median	20.04 DU
25-75 percentile range	15.04 DU to 21.78 DU
10-90 percentile range	12.61 DU to 25.49 DU
min-max range	8.90 DU to 26.89 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	52.96 %

UMCP$_\gamma$



longitude, latitude	-76.942519, 38.990906
n, nWI, nSP, nSU, nAU	87, 19, 17, 27, 24
max Pan.	16
max Pan.loc. / footp.	6
CC	0.98
slope	1.04 ± 0.00
offset	-15.16 DU ± 0.48 DU
std	7.55 DU
mean	4.17 DU
median	4.42 DU
25-75 percentile range	-0.69 DU to 10.17 DU
10-90 percentile range	-4.30 DU to 13.36 DU
min-max range	-17.59 DU to 24.92 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	56.44 %

USNA$_\gamma$



longitude, latitude	-76.479558, 38.983673
n, nWI, nSP, nSU, nAU	66, 16, 12, 22, 16
max Pan.	20
max Pan.loc. / footp.	6
CC	0.98
slope	1.02 ± 0.00
offset	-9.96 DU ± 0.55 DU
std	7.71 DU
mean	4.05 DU
median	4.86 DU
25-75 percentile range	-1.28 DU to 9.50 DU
10-90 percentile range	-3.19 DU to 14.65 DU
min-max range	-17.59 DU to 24.92 DU
SCIAMACHY CF	100.00 %
Pandora DQ0	26.32 %

3.2.4 Conclusions of Ozone Validation

Table 7 lists n, MEDD, $\text{MEDD}_{\text{TCORR}}$, CC and CC_{TCORR} for each validation location. $\text{MEDD}_{\text{TCORR}}$ (CC_{TCORR}) is the corrected MEDD (CC), where a very rudimentary method to remove the effect of the effective O_3 temperature was applied based on the findings from section 3.2.2. We add (remove) 7 DU to (from) the Pandora TO3 summer (winter) values before forming the median difference and correlation coefficient. The spring and autumn values remain untouched, since we assume the constant gas temperature to be suitable throughout these months. The $\text{MEDD}_{\text{TCORR}}$ and CC_{TCORR} in Table 7 (fourth and sixth column) clearly show an improvement in the comparison quality.

Location	**n**	**MEDD**	**$\mathrm{MEDD}_{\mathrm{TCORR}}$**	**CC**	**$\mathrm{CC}_{\mathrm{TCORR}}$**
Aldino_γ	6	14.94 DU	7.94 DU	0.96	0.96
$\mathrm{Beltsville}_\gamma$	85	4.45 DU	4.02 DU	0.97	0.98
Busan	5	4.51 DU	4.51 DU	0.97	0.97
Cabauw_α	13	0.55 DU	-6.45 DU	0.89	0.89
$\mathrm{Edgewood}_\gamma$	9	12.37 DU	5.59 DU	0.97	0.98
Essex_γ	10	18.40 DU	12.47 DU	0.94	0.96
FMI	18	4.23 DU	4.23 DU	0.99	0.99
FSU_γ	4	8.37 DU	6.58 DU	0.99	1.00
GSFC_γ	106	4.94 DU	3.94 DU	0.97	0.98
IZO	17	4.40 DU	5.87 DU	0.99	0.98
Langley	95	2.09 DU	3.49 DU	0.97	0.98
NASAHQ_γ	106	4.94 DU	3.94 DU	0.97	0.98
$\mathrm{Oldtown}_\gamma$	17	19.87 DU	12.87 DU	0.95	0.95
$\mathrm{Padonia}_\gamma$	10	13.43 DU	7.16 DU	0.94	0.95
SERC_γ	65	5.54 DU	4.02 DU	0.97	0.98
UMBC_γ	14	20.04 DU	13.64 DU	0.89	0.89
UMCP_γ	87	4.42 DU	4.02 DU	0.98	0.98
USNA_γ	66	4.86 DU	3.81 DU	0.98	0.98

Table 7: n, MEDD, $\mathrm{MEDD}_{\mathrm{TCORR}}$, CC, and $\mathrm{CC}_{\mathrm{TCORR}}$ for each location in this validation study. $\mathrm{MEDD}_{\mathrm{TCORR}}$ is calculated by reducing (enhancing) the winter (summmer) values of the Pandora TO3 values by 7 DU before forming the median difference. The same for $\mathrm{CC}_{\mathrm{TCORR}}$.

The majority of the validated locations offer multiple Pandora units for the SCIAMACHY validation (marked by γ). This situation occurs mainly in the Baltimore-Washington area, as almost every SCIAMACHY footprint embraces GSFC_γ. This makes $\mathrm{Beltsville}_\gamma$, NASAHQ_γ, UMCP_γ and USNA_γ mainly driven by GSFC_γ data and therefore give very similar comparison results.

All stations have in common a 100 % admittance of SCIAMACHY data, i.e. all SCIAMACHY data pass the SciaCF = 0.5 threshold. This is due to the Pandora DQ0 data quality filter, which filters cloudy conditions and automatically excludes high SciaCF cloud fractions. We allow all Pandora data $\pm$ PanAVG = 200 min to be averaged around the SCIAMACHY overpass time. Assuming average high altitude wind speeds in mid latitudes ($>$ 10 $\mathrm{ms}^{-}1$), an air parcel is able to cross the SCIAMACHY footprint within 200 minutes. That is if no Pandora data is allowed within 200 minutes the cloud cover is supposed to be larger than 4/8.

The MEDD range from 0 to 20 DU and the $\mathrm{MEDD}_{\mathrm{TCORR}}$ from -7 to 13 DU. Excluding Cabauw, all CC are above 0.95 and the $\mathrm{MEDD}_{\mathrm{TCORR}}$ are between 3 and 13 DU. On average the SCIAMACHY TO3 are 8 $\pm$ 5 DU (about 1-4 %) higher than the Pandora TO3. This bias can have several reasons. From the Pandora side the most likely explanation is a too low calibration standard. Remember, that the Pandora data are not referenced to any standard instrument, but use a calibration purely from the laboratory. More important than the bias is the excellent homogeneity of the Pandora data, which is the main goal of a ground based network [RD04].

Cabauw (only summer data) is different from all other locations with CC = 0.89 and a $\mathrm{MEDD}_{\mathrm{TCORR}}$ = -6 DU, i.e. SCIAMACHY about 2 % lower than Pandora. The negative bias could be related to the geolocation of Cabauw_α. At 50°N, the average summer temperatures are expected to be lower compared to e.g. GSFC_γ. The underestimation in temperature would lead to an underestimation in TO3 for the Pandora retrieval.

An improvement of the TO3 validation study could be obtained, if the Pandora TO3 algorithm was able to retrieve the effective effective O_3 temperature instead of using a constant temperature. Then one could possibly also determine the satellite error in TO3 due to the use of a climatology for the effective O_3 temperature.

4 Nitrogen Dioxide Validation

As for O_3, we based the NO_2 validation on the SCIAMACHY vertical column amounts provided by ESA and applied the validation strategy outlined in section 4 of RD06. Figure 5 shows a comparison of TNO2 from Pandora and SCIAMACHY at GSFC using the ESA-provided data in the left panel and the SCIAMACHY retrievals from `http://www.temis.nl/airpollution/no2col/data/scia/overpass/Goddard_sciano2.dat` in the right panel. The right panel is identical to figure 2 of RD06, showing an excellent average agreement between ground and satellite observations. The satellite data in the left panel underestimate the column amounts by approximately a factor of 2. The SCIAMACHY(ESA) data scaled by this factor are shown in gray color. The same large discrepancy in the provided SCIAMACHY NO_2 columns is seen at each single Pandora location. This makes any further analysis meaningless.

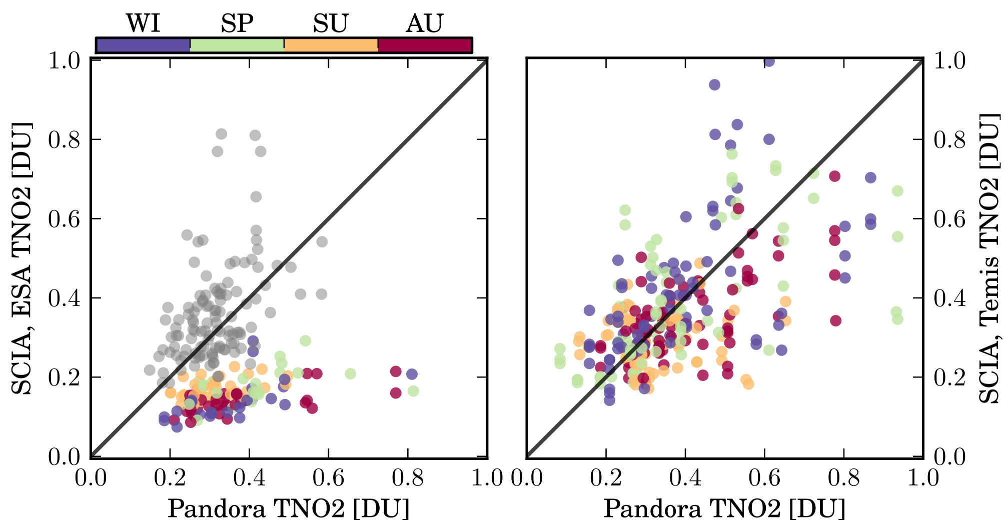


Figure 5: TNO2 scatter plot for SCIAMACHY and Pandora data from $GSFC_\gamma$. The colored data in the left panel show SCIAMACHY TNO2 extracted from SCIAMACHY level 2 files, in the right panel data from Temis. The SCIAMACHY TNO2 underestimate the Pandora TNO2 by approximately a factor of 2. The gray data are the SCIAMACHY data multiplied by 2. The color code reflects the different seasons.

4.1 Conclusions of Nitrogen Dioxide Validaton

The SCIAMACHY TNO2 provided by ESA are largely underestimating the true NO_2 columns. This is not a problem of data extraction and interpretation from the provided files, since we have checked carefully whether the correct columns were used or any scaling or offset factors were not applied. The underestimation of approximately a factor of 2 is due to the SCIAMACHY retrievals of TNO2 for polluted areas, which does not use a proper tropospheric air mass factor for such situations.. Therefore we did not proceed with a detailed validation analysis as we did for O_3. This is unfortunate since for TNO2 we had expected interesting results from a sensitivity study (as in section 3.1 for TO3) and also were looking forward to test the climatological correction outlined in the validation strategy (see section 4 of RD06).