Maximizing the Use of Pandora Data for Scientific Applications

Prajjwal Rawat[1], James H. Crawford[1], Katherine R. Travis[1], Laura M. Judd[1], Mary Angelique G. Demetillo[1], Lukas C. Valin[2], James J. Szykman[2], Andrew Whitehill[2], Eric Baumann[2], Thomas F. Hanisco[3]

[1]NASA Langley Research Center, Hampton, VA, 23681, United States
[2]Environmental Protection Agency Office of Research and Development, Research Triangle Park, NC, 27709, US
[3]NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

Correspondence to: Prajjwal Rawat (prajjwal.rawat@nasa.gov) and James H. Crawford (james.h.crawford@nasa.gov)

Supplementary materials

Supplementary text

S1.1 Pandora Instruments, data retrieval and data quality flagging: A brief review

The first operational software of Pandora was called Pan Software Suite (PSS). In 2018, the PGN team introduced the Blick Software Suite (BSS) as an evolution of PSS, incorporating various new software features (Cede et al., 2021). The main changes from PSS to BSS were real-time data, uniform file naming, quality assurance of raw and retrieved data, stray light correction, and the inclusion of sky-scan (SS) column retrievals. Currently, PGN utilizes BSSv1.8 software, released in 2020 and built upon different software iterations, including PSS and BSSv1.7. The new processor version 1.8 produces improved total column amounts using better estimations for the effective temperature and the air mass factors. Also, it adds new PGN products, surface concentrations and tropospheric profiles of NO_2 and formaldehyde (HCHO) based on sky observations and includes random and common uncertainties (Cede et al., 2021, 2023).

Regarding the hardware aspects, a major modification, among various design changes, was the replacement of the front window of the telescope assembly with a window containing an anti-reflective coating (ARC) to minimize non-atmospheric spectral signal in the direct-sun (DS) observations, which is called "unwanted spectral signal" (USS) and degrades spectral fitting. This modification effectively mitigates etaloning interference and improves the ability to retrieve HCHO columns from the Pandora DS measurements (Cede et al., 2022). However, over time, the anti-reflective coating (ARC) tended to deteriorate, and the ultimate solution to reduce etaloning was achieved using the wedged window. The results showed that the weighted spectral fit RMS for the wedged windows were as good as for the case of no window hence preferred the wedged windows as the right optical choice for reducing unwanted spectral signal (LuftBlick_FRM4AQ, Fiducial Reference Measurements for Air Quality). Additionally, in 2019, it was observed that components made of Delrin in the head sensor were emitting formaldehyde (HCHO) gasses; subsequently, SciGlob replaced materials for these pieces from Delrin to black Nylon (Spinei et al., 2021).

In DS mode, Pandora alignment is configured such that most of the forward scattered diffuse light is sampled within a narrow field of view (FOV). Specifically, solar radiation is recorded by the Pandora front-end optics with a FOV of 2.5° for DS observations using a diffuser and a FOV of 1.5° for SS observations without a diffuser. In the DS diffuser mode, most of the light is collected within an angle of 0.5°. The spectrometer incorporates 2048×64 pixels back-thinned Hamamatsu charge-coupled device (CCD) with a 50µm entrance slit and a grating with 1200 lines mm^{-1} (Cede et al., 2022). After radiance correction, the slant column densities (SCDs) are estimated following DOAS analysis and using a static reference spectrum (taken months apart) from the modified Langley extrapolation method (MLE) (Herman et al. 2009) or extended MLE methods (in e-MLE the tropospheric SCDs are further subtracted in direct sun using SS SCDs) (Cede et al., 2021). The successive retrieval of total vertical column densities (VCD) is performed by dividing the SCD by geometrical AMF. The primary source of uncertainty in the DS retrieval comes from the AMF calculation error, reference spectrum, and DOAS fitting error. The reported accuracy of the DS VCD of NO_2 is 0.05 Dobson unit (1 DU = 2.7×10^{15} molecules cm^{-2}) with a precision of 0.002 DU (5.3×10^{13} molecules cm^{-2}) (Harman et al., 2009; Cede, 2021), while for HCHO a smaller statistical error up to 6% and notable systematic error up to 26% are reported (Spinei et al. 2018; 2021).

In SS mode, Pandora records scattered light at various zenith angles (0, 60, 75, 88, and 89 degrees) while maintaining a fixed azimuth direction, mostly oriented northward. The Pandora SS retrieval is a fast analytical method (NASA algorithm) based on rigorous background radiative transfer modeling (Cede et al., 2021; Frieß et al., 2019; Tirpitz et al., 2021), with coherent reference spectrum from zenith for individual measurements. The NASA algorithm is used in the CINDI-2 intercomparisons study utilizing the harmonized or median SCDs of participating instruments. It reported nominal bias and strong correlation for both the HCHO and NO_2 retrieval by the NASA algorithm during clear and cloudy days, which further improved when using its own SCDs rather than the median values. Similarly, Frieß et al., (2019) also applied different MAX-DOAS retrieval algorithms, including the NASA algorithm, to the simulated SCDs with different atmospheric scenarios and reported reasonable output with minimal computational cost for the NASA analytical algorithm. In brief, the fast NASA algorithm relies primarily on a pure Rayleigh atmosphere, with the optical light path (AMF) approximated from the typical ratio of oxygen dimer (O_2-O_2) SCDs in DOAS fitting with a known O_2-O_2 vertical column from climatology, as explained by Cede et al., (2021). Generally, the derived tropospheric column is a scattered response of photons, mainly in the lower troposphere up to 4 km, and the weighted profiles are columns at various heights up to 4-5 km, but most typically in the range of 2-4 km. The exact uncertainty budget estimation for SS measurements is very challenging, and Pandora reports the measurement uncertainty for each retrieved product. However, SS has minimal reference spectrum uncertainty due to coherent zenith reference for each retrieval, unlike DS having a nearly fixed reference spectra and vice-versa for AMF.

Additionally, the quality assurance of Pandora L2 data in BSSv1.8 is structured on the three-step processes to assign one of nine possible quality flags. In the first step, various retrieval parameters, i.e., uncertainties, wrms (normalized rms of the fitting residuals weighted with the independent uncertainty), atmospheric variability (a measure of radiance variance of the detector), wavelength shift, and AMF, are set to some threshold values to categorize the data in high, medium, and low quality. The threshold values are estimated using the Gaussian mixture regression model and described in the Fiducial Reference Measurements for Air Quality report of Luftblick for each retrieved parameter (Gebetsberger et al., 2023b). This step provides quality flags as high (10), medium (11), and low (12) quality data without quality assurance. In the second quality assurance step, the time series fluctuation of the multiple quality indicators of the same data product is studied for a more extended period and sudden relative changes or drifts are used to categorize the data in high (0), medium (1), and low (2) quality. However, these steps do not account for various other potential sources of errors in the data, such as misalignment of instruments, telescope collimator errors, or unknown spectral features. To address such issues, a third quality assurance procedure is implemented based on time series fluctuation of multiple parameters of the multiple data products and assign invalid/unusable data flags as high unusable (20), medium unusable (21), and low unusable (22), data if anomalous behavior is observed. This is discussed in more detail by Gebetsberger et al., (2023b).

S1.2 Pandora Observational Schedule

The observational schedule of Pandora is a sequence of routines to be executed by the Pandora instruments for different modes of observation. The standard schedules are stored in the BSS routine library and described by two-letter codes. Below are the few codes discussed in the present manuscript, and the details of all possible routines can be found on page no. 63 of Cede et al., (2021). We also give the calculation to convert t_{eff} (Column 3: Effective duration of measurement [s]) to total measurement duration (Δt) for SS routines, which only provide this value for the maximum zenith angle in Pandora data files.

1. **Elevation scan routines E ***

EK = Detailed elevation scan at standard azimuth without filter; the standard azimuth is the one given in the BlickO Instrument Configuration File, entry "Standard azimuth for elevation scans [deg]". All elevation scan routines also use entry "Maximum unobstructed pointing zenith angle at standard azimuth [deg]" from the configuration file. This routine has the following angles [0,60,70,75,"MAX-9","MAX-8","MAX-7","MAX-6","MAX-5","MAX-4","MAX-3","MAX-2","MAX-1","MAX","MAX-1","MAX-2","MAX-3","MAX-4","MAX-5","MAX-6","MAX-7","MAX-8","MAX-9",75,70,60,0], where for a flat horizon MAX angle is set to 89.0deg. It could also be larger or smaller than 89.0deg, depending on the horizon of the location at the standard azimuth. $\Delta t = 26 * t_{eff} /2 + t_{eff}$. EL = Detailed elevation scan at standard azimuth with U340. This routine has the following angles [0,40,50,60,70,75,80,82,85,"MAX-2","MAX-1","MAX","MAX-1","MAX-2",85,82,80,75,70,60,50,40,0], where for a flat horizon MAX angle is set to 89.0 deg. It could also be larger or smaller than 89.0 deg, depending on the horizon of the location at the standard azimuth. $\Delta t = 22 * t_{eff} /2 + T_{eff}$.

EO = Quick elevation scan (5 zenith angles) at standard azimuth without filter. This routine has the following angles [0,60,75,"MAX-1","MAX","MAX-1",75,60,0], where for a flat horizon

MAX angle is set to 89.0deg. It could also be larger or smaller than 89.0 deg, depending on the horizon of the location at the standard azimuth. $\Delta t = 8 * t_{eff} / 2 + t_{eff}$.

EU = Quick elevation (5 zenith angles) at standard azimuth with U340. This routine has the following angles [0,60,75,"MAX-1","MAX","MAX-1",75,60,0], where for a flat horizon MAX angle is set to 89.0 deg. It could also be larger or smaller than 89.0 deg, depending on the horizon of the location at the standard azimuth. $\Delta t = 8 * t_{eff} / 2 + t_{eff}$.

Find routines F*

FS = Find Sun; long sun search for all spectrometers, saves final figure and averaged data, but not spectral data

FD = FS, but with diffuser in the optical path

FO = FD but scanning only around the center of the FOV with longer measurement duration. This is not used to "find" the sun.

FP = FU but scanning only around the center of the FOV with longer measurement duration. This is not used to "find" the sun.

FU = FS, but with U340 plus diffuser combination in the optical path

FW = FS with spectra, but showing the results in 4 wavelength regions

2. **Direct Sun routines S***

SQ = Quick direct-Sun without filter

SS = Quick direct-Sun with U340

3. **Zenith sky routines Z***

ZO = Zenith-Sky without filter

ZU = Zenith-Sky with U340

4. ** manual routine

Supplementary Table

Table S1. Pandora L2 file naming for the sky-scan and direc-sun measurements and corresponding column number for vertical column, independent uncertainty, wrms, MHzD, L2 quality flags, and effective measurement time.

Species	HCHO		NO_2	
Mode	**Sky-scan**	**Direct-sun**	**Sky-scan**	**Direct-sun**
Pandora L2 file naming	*rfuh5p1-8	*rfus5p1-8	*rnvh3p1-8	*rnvs3p1-8
Column # for vertical column	49	39	62	39
Column # for independent uncertainty	50	40	63	40
Column # for wrms	11	9	11	9
Column # for MHzD	51	N/A	64	N/A
Column # for L2 Quality flag	42	36	53	36
Column # for effective measurement time	3	3	3	3

Table S2. Name and location of Pandora sites used in this study.

Pandora #	Site Name	Lat, Lon (deg)	PIs
61	AldineTX	29.9011, -95.3262	Tom Hanisco
25	HoustonTX	29.7200, -95.3400	Jimmy Flynn
58	LaPorteTX	29.6721, -95.0647	Tom Hanisco
35	ManhattanNY	40.8153, -73.9505	Maria Tzortziou
38	BayonneNJ	40.6703, -74.1261	Maria Tzortziou
180	BronxNY	40.8679, -73.8781	Luke Vailn
177	WestportCT	41.1183, -73.3367	Luke Vailn
69	NewBrunswick	40.4622, -74.4294	Nader Abuhassan
64	NewHavenCT	41.3014, -72.9029	Nader Abuhassan
55	QueensNY	40.7361, -73.8215	Jim Szykman
134	BristolPA	40.1074, -74.8824	Luke Vailn
166	Philadelphia	39.9919, -75.0811	Luke Vailn
140	WashingtonDC	38.9218, -77.0124	Jim Szykman
156	HamptonVA-HU	37.0203, -76.3366	John Anderson
145	Toronto-S	43.7843, -79.1874	Vitali Fioletov

Supplementary Figures

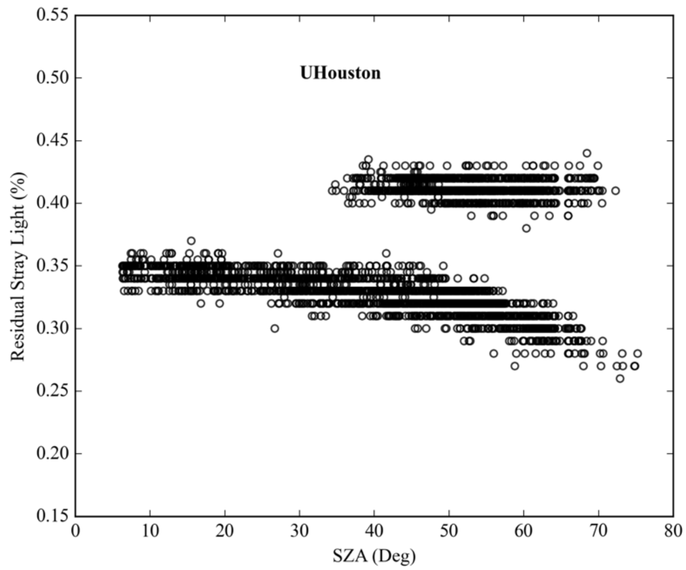

Figure S1. Residual stray light along solar zenith angle over University of Houston. Analysis was done for other sites as well and similar pattern was found for most of the sites.

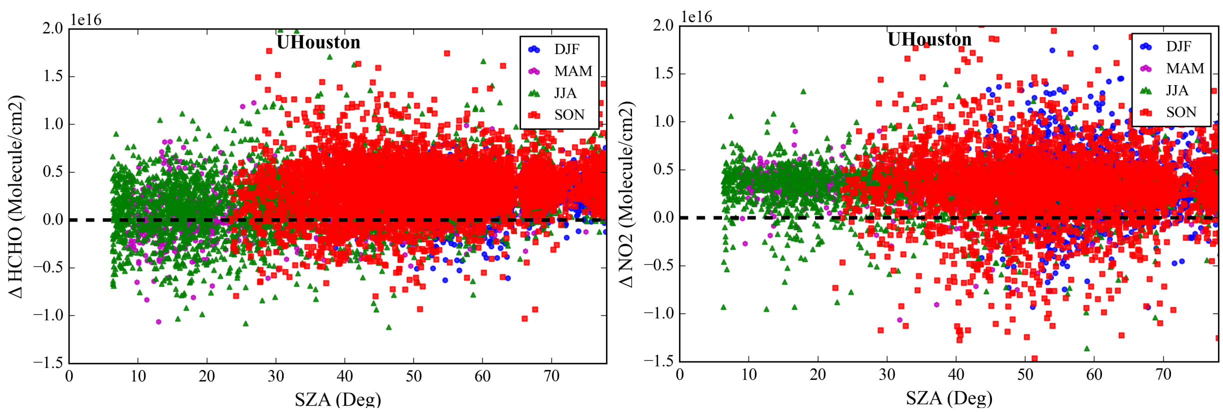

Figure S2. HCHO (left) and NO_2 (right) column differences for different seasons (DJF is winter, MAM is spring, JJA is summer, and SON is autumn) along varying SZA at the University of Houston between DS and SS. Other sites are also studied, and similar results were noted.

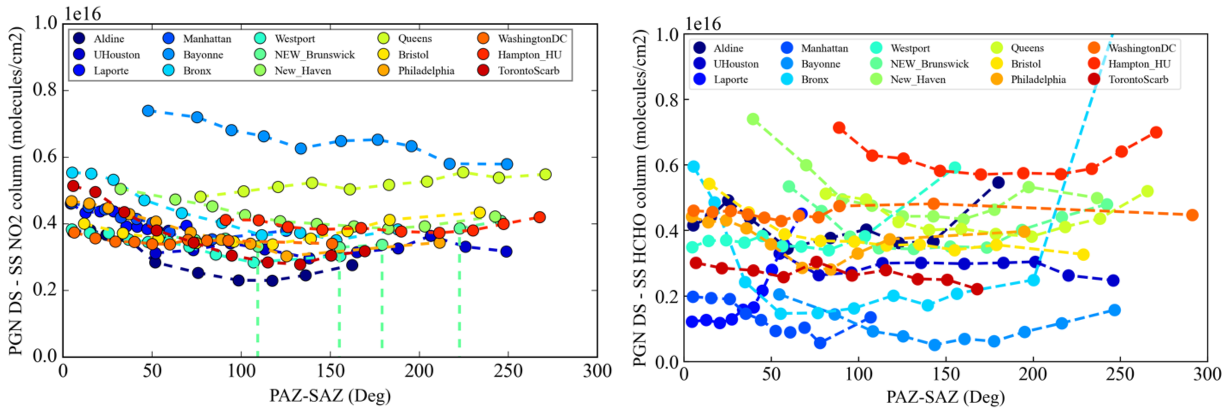

Figure S3. HCHO and NO_2 column differences along differences in pointing azimuth and solar azimuth angles (PAZ-SAZ) over 15 Pandora sites between DS and SS measurements.