Comparison of gas columns retrieved by TEMPO and Pandora – ASDC tools, results, and new development

Alexander Radkevich[1,3], Daniel Kaufman[2,3]

[1] Adnet Systems, [2] Booz-Allen and Hamilton, Inc., [3] Atmospheric Science Data Center, NASA Langley Research Center

Introduction

Monitoring emissions of nitrogen dioxide and formaldehyde is crucial for understanding atmospheric composition and its impacts on air quality and climate. This study uses retrievals of gas columns from space- and ground-based sensors with the main goals of (1) presenting programmatic tools comparing those retrievals and (2) evaluating the accuracy of retrievals by Tropospheric Emissions: Monitoring of Pollution (TEMPO) [1], a new mission with data made publicly available in May 2024 via the U.S. National Aeronautics and Space Administration's (NASA's) Earthdata.

Instruments and data

The Tropospheric Emissions: Monitoring of Pollution (TEMPO) instrument is a grating spectrometer, sensitive to visible and ultraviolet light within spectral ranges of 290-490 nm and 540-740 nm and 0.6 nm spectral resolution. The TEMPO instrument is aboard the Earth-facing side of a commercial telecommunications satellite (Intelsat 40e) in geostationary orbit over 91° W longitude. This positioning allows TEMPO to maintain a continuous view of North America and TEMPO's light-collecting mirror can complete hourly east-to-west scans of the field of regard during daylight hours. High resolution measurements of radiance reflected by the Earth's back to the instrument's detectors enable retrievals of atmospheric columns of ozone, nitrogen dioxide, and formaldehyde, each of which play important roles in the chemical dynamics of Earth's atmosphere [2]. The retrieved columns are stored as Level 2 (used in this presentation) and 3 granules.

Mission objectives for TEMPO involve understanding the dynamics of air quality, pollution sources, and their impact on climate change. By providing near real-time data and comprehensive atmospheric composition measurements, TEMPO will assist scientists in studying pollution patterns, evaluating the efficacy of environmental policies, and predicting future trends in air quality.

In 2005 NASA initiated an effort at Goddard Space Flight Center (GFSC) to develop a cost-effective, compact, easy to deploy, ground-based, passive remote sensing spectrometer system capable of performing sun, moon, and sky observations called "Pandora". Pandora has continuously evolved since then with support from NASA and the European Space Agency (ESA) and has been manufactured by SciGlob LLC, Elkridge, MD, USA since 2010. Soon thereafter Pandoras were distributed around the globe to form a *validation network*. In 2013 ESA joined NASA in funding this development through prime contractor LuftBlick OG, Innsbruck, Austria. Since 2018, this network is called "Pandonia Global Network" (PGN) and aims to ensure systematic processing and dissemination of the data to the greater global community in support of air quality monitoring and satellite validation. The PGN is a federation carried out jointly by NASA and ESA as part of their "Joint Program Planning Group Subgroup" on calibration and validation and field activities [3]. Pandora data include retrievals of NO_2 and HCHO total and tropospheric columns. This study uses Level 2 versions of these data products.

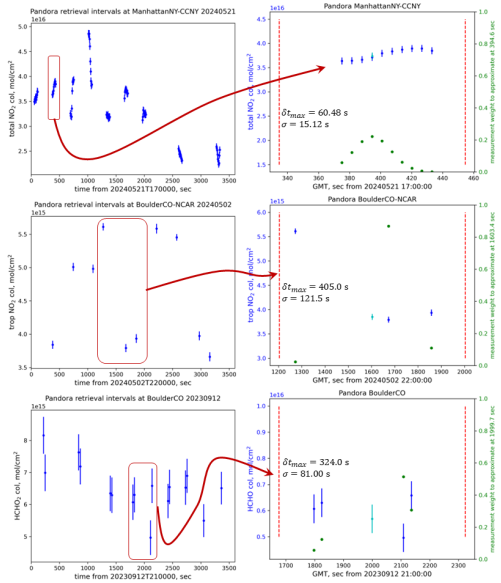

Figure 1. Hourly timeseries of Pandora NO_2 and HCHO columns (left, blue) and those column within the running window (right, blue) used to compute average column at a random time of interest (cyan), see Eq. (1). Measurement weights are shown in green.

Temporal features of Pandora retrievals

Temporal resolution of TEMPO varies with the rate of scanning, e.g., there are granules with timestamps 20230827T214434Z_S028G01and 20230827T215527Z_S029G01 covering the same part of North America within 11 minutes; however, nominal temporal resolution is 1 hour, e.g., granules 20240827T202615Z_S011G08 and 20240827T212615Z_S012G08 also cover the same area. From Pandora, there may be tens of measurements per hour. The frequency of measurements depends on the product. Left panels of Fig. 1 show measurements of different trace gases within an hour interval. While direct comparison of non-simultaneous measurements from TEMPO and Pandora gives some qualitative understanding of their differences, it is highly desirable to bring them to the same times in a plausibly realistic way accounting for the volatile nature of the gas columns. This study proposes the use of "running window" averages with Gaussian-like weights as described in Eq. (1):

$$x_{int}(t) = \sum_m x(t_m) w(t_m, t), \quad w(t_m, t) = \exp\left[-\frac{(t-t_m)^2}{2\sigma^2}\right] \Big/ \left(\sum_m \exp\left[-\frac{(t-t_m)^2}{2\sigma^2}\right]\right)^{-1}, \quad |t - t_m| \le \delta t_{max} \quad (1)$$

where sums are taken over Pandora measurement times t_m within δt_{max} from the target time of interest t. The temporal features of Pandora retrievals are accounted for by selection of product-dependent values of σ and δt_{max}, see Fig. 1.

Co-locating Pandora stations and TEMPO pixels

Trace gases columns may vary by an order of magnitude between neighboring pixels, see Table 1. For this reason, instead of using a Pandora station's nearest neighbor, this study averages TEMPO retrievals from 4 pixels surrounding it, Fig. 2.

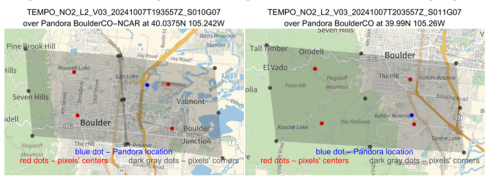

Figure 2. Spatial extents of TEMPO L2 pixels with the quadrilateral of the centers containing the point of interest

Table 1. TEMPO HCHO pixels at 20241007T193557Z_S010G07 at BoulderCO-NCAR (40.0375°N, 105.242°W)

line	pixel	latitude, deg N	longitude, deg W	HCHO tot col, 10^{15} mol/cm^2	HCHO tot col unc, 10^{15} mol/cm^2	QF
19	691	40.038895	105.229225	3.24	4.05	0
19	692	40.018490	105.224243	**-1.81**	4.74	**0**
20	691	40.040039	105.287163	15.69	4.65	0
20	692	40.020042	105.282364	6.17	5.05	0

Python notebooks' structure and output

ASDC has created python notebooks [4] comparing O_3 total, NO_2 total, HCHO total, and NO_2 tropospheric columns from corresponding TEMPO products against Pandora retrievals. Besides importing necessary libraries and setting up access to NASA's Earthdata, the notebooks perform the following steps:

1) Available Pandora stations are list, and the user is prompted to select one of them;
2) Appropriate Pandora data product is downloaded;
3) User is prompted to enter a timeframe of interest;
4) Pandora data file is read within the timeframe to form a timeseries of columnar retrievals with high quality flags;
5) Appropriate TEMPO granules falling into the timeframe and covering the Pandora location are found and downloaded by means of the earthaccess library;
6) The TEMPO granules are scanned for pixels surrounding the Pandora station, see Fig. 2. Retrievals from those pixels are averaged while accounting for the quality flags (retrievals from low and bad quality pixels are discarded); results of this step are written into an ASCII file to avoid time-consuming computations and to facilitate quality control;
7) Timeseries from both instruments are plotted and saved;
8) Pandora retrieved columns are averaged to the TEMPO times of observation (please bear in mind that some TEMPO times may not have corresponding Pandora averages due to limitations of Eq. (1));
9) Scatter plots showing Pandora vs TEMPO and a regression analysis are created and saved.

Note 1: TEMPO_NO2_L2_V03 granules do not have a variable for total NO_2 column in their /product group. Instead, the User Guide [5] recommends adding up vertical_column_stratosphere and vertical_column_troposphere datasets from that group. Moreover, the use of /support_data/vertical_column_total variable is depreciated by the User Guide. A potential complication with this approach is that uncertainty of vertical_column_stratosphere is not provided, so that uncertainty of the total column is that of tropospheric column.

Note 2: TEMPO science team recommendations on the use of quality flags also can be found in the User Guide [5].

Note 3: Since TEMPO retrievals can be negative, see Table 1, three sets of high quality (HQ) retrievals are accumulated and analyzed: (1) all HQ, (2) positive only HQ, and (3) HQ retrieval greater than corresponding uncertainty. The last one is accumulated to check whether removal of "white noise" can improve match between the instruments.

Future updates and improvement

The current Python notebooks enabling interactive comparison of TEMPO and Pandora retrievals of trace gases columns were created based on V03 of the TEMPO products. The next version of the product is available from September 16, 2025. The TEMPO Science Team is currently performing backward re-processing of the previously acquired observations into version V04 products. Both versions are currently available via the NASA Earthdata site.

While modification of the notebooks to accommodate the new versions is straightforward, ASDC is currently working on another known issue with the current notebooks - they download whole granules of the products while a small portion of available variables is needed and a tiny geographic subset of a granule. Downloading of what is needed instead of what is available can be achieved by using the L2 subsetter [6]. Though subsetting on the data provider side takes some time, significant reduction of downloading time ensures overall acceleration of the statistics accumulation. The new versions of the notebooks are currently under development.

Results: total nitrogen dioxide column

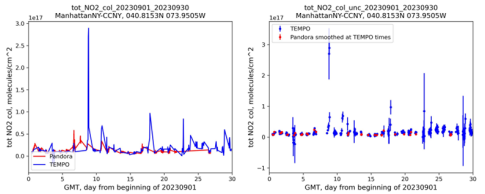

Figure 3. Comparison of TEMPO and Pandora timeseries of NO_2 total column at original times (left) and with Pandora measurements averaged to coincide with TEMPO times (right). These two figures permit a qualitative sense of the comparison, but quantitative comparisons are aided by the scatter plots and analyses below.

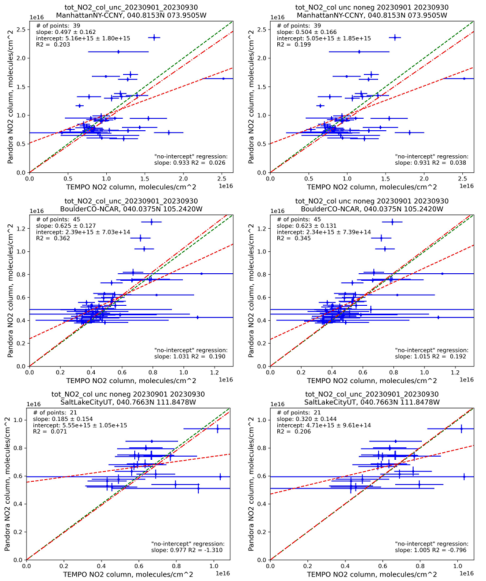

Figure 4. Scatter plots of nitrogen dioxide total columns retrieved by TEMPO and Pandora (official data from all 3 stations), showing all high quality (left) and positive-only, high quality (right) retrievals. Green dashed, red dash-dot, and red dashed lines indicate 1:1 reference, "no-intercept" regression, and general linear regression.

References

1. https://tempo.si.edu/overview.html
2. https://asdc.larc.nasa.gov/project/TEMPO
3. https://www.pandonia-global-network.org/home/about/about-pgn/
4. https://github.com/nasa/ASDC_Data_and_User_Services/tree/main/TEMPO/L2_validation_codes
5. https://asdc.larc.nasa.gov/documents/tempo/guide/TEMPO_Level-2-3_trace_gas_clouds_user_guide_V2.1.pdf
6. https://podaac.github.io/l2ss-py/

Results: total formaldehyde column

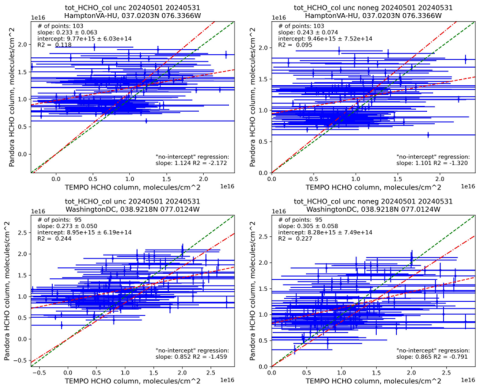

Figure 6. The same as Fig. 4, but for formaldehyde total columns, Pandora data from both stations are official.

Results: tropospheric nitrogen dioxide column

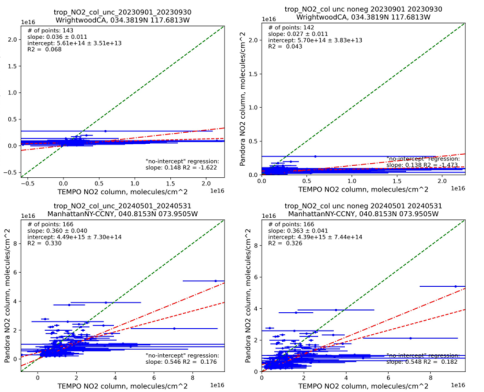

Figure 7. The same as Fig. 4, but for nitrogen dioxide tropospheric columns. Official Pandora data from both stations.

Conclusion

Python notebooks enabling interactive comparison of TEMPO and Pandora retrievals of trace gases columns were created and made available via the NASA Atmospheric Science Data Center (ASDC) GitHub page.

Limited results included in this presentation show reasonable match between total nitrogen dioxide and total formaldehyde columns retrieved by TEMPO and Pandora. However, the non-trivial intercept of the general linear regression and low coefficient of determination confirm the visually apparent shapelessness of clouds of data points in the retrieval scatter plots, i.e., these retrievals from the two instruments are poorly correlated. Removal of physically meaningless negative retrievals from consideration slightly improves regression statistics by reducing intercepts and moving slopes closer to 1, but this does not eliminate discrepancies. It was also found that discrepancies between all trace gas (especially formaldehyde total) columns from Pandora and TEMPO are site-dependent. NO_2 tropospheric columns from TEMPO and PANDORA are barely correlated with noticeable overestimation by TEMPO. Further optimization of the notebooks with the use of data sub-setter constitutes future work.