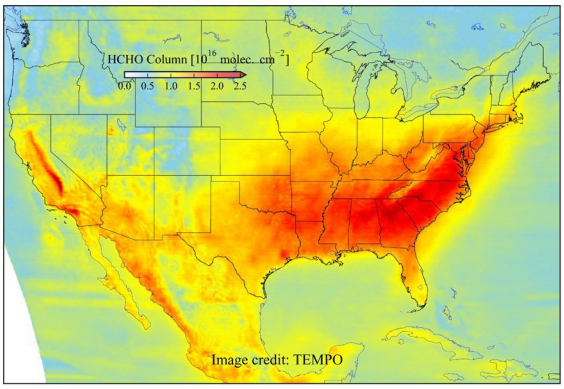
EPA/600/R-25/246 | September 2025

United States
Environmental Protection
Agency

National Aeronautics and
Space Administration

Smithsonian



TROPOSPHERIC EMISSIONS: MONITORING OF POLLUTION (TEMPO) PROJECT

Validation and Quality Assessment of the TEMPO Level-2 Trace Gas Products

DOI: 10.5067/doc/tempo/L2TraceGasValidationAssessment

EPA/600/R-25/246 | September 2025

United States
Environmental Protection
Agency

TROPOSPHERIC EMISSIONS:
MONITORING OF POLLUTION (TEMPO) PROJECT

Validation and Quality Assessment of the
TEMPO Level-2 Trace Gas Products

DOI: 10.5067/doc/tempo/L2TraceGasValidationAssessment

EPA/600/R-25/246 | September 2025

United States
Environmental Protection
Agency

TROPOSPHERIC EMISSIONS:
MONITORING OF POLLUTION (TEMPO) PROJECT

Validation and Quality Assessment of the
TEMPO Level-2 Trace Gas Products

Prepared by the TEMPO Validation Team and TEMPO Ad-hoc Validation Working Group

Working Group Leads: James Szykman (U.S. EPA), R. Bradley Pierce (UW-Madison), Barron Henderson (U.S. EPA) and Laura Judd (NASA)

DOI: 10.5067/doc/tempo/L2TraceGasValidationAssessment

DISCLAIMER

This document has been reviewed in accordance with the U.S. Environmental Protection Agency policy and approved for publication. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government, the U.S. Environmental Protection Agency, or the Smithsonian Astrophysical Observatory.

This work is conducted under the U.S. Environmental Protection Agency's Quality Assurance (QA) program for environmental information, with an approved Quality Assurance Project Plan: "Routine collection of solar and sky ultraviolet and visible radiance and delivery of raw counts data to the Pandonia Global Network", J-AESMD-0033858-QP-1-0.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

AUTHORS, CONTRIBUTORS, AND REVIEWERS

Scientific and Technical Direction

Mr. James Szykman - Center for Environmental Measurement and Modeling, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC
Dr. Brad Pierce - Space Science and Engineering Center, University of Wisconsin-Madison, Madison, WI
Dr. Barron Henderson - Center for Environmental Measurement and Modeling, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC
Dr. Laura Judd – Science Directorate, NASA Langley Research Center, Hampton, VA

Authors

Section 1 Introduction

James Szykman (EPA), Brad Pierce (Space Science and Engineering Center, Univ. of Wisc.), Barron Henderson (EPA), Laura Judd (NASA Langley Research Center)

Section 2 Instrument, Level 2 Science Product Description and associated Product Maturity Level Summary

James Szykman (EPA), Brad Pierce (Space Science and Engineering Center, Univ. of Wisc.), Laura Judd (NASA Langley Research Center). Xiong Liu (SAO)

Section 3 Validation Results

Section 3.1 Status Summary of Level 1B Radiance and Irradiance

Authored by: Heesung Chong, Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO-TEMPO)

Section 3.2 Level 2 Science Product Validation and Evaluation TEMPO_NO2_L2_V03 - TEMPO NO2 tropospheric and stratospheric columns V03

Section 3.2.1 TropNO2 Inter-Comparison between TEMPO and TROPOMI

Authored by: **(Shobha Kondragunta)** Aerosol and Atmospheric Composition Group at NOAA/STAR

Section 3.2.2 TropNO2 Inter-Comparison Between TEMPO and TROPOMI

Authored by: **Barron Henderson**, Lukas Valin, and Jim Szykman (EPA); Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO-TEMPO)

Section 3.2.3 TotNO2 Inter-Comparison Between TEMPO and the Pandonia Global Network

Authored by: **Ivan Ortega**, Sara Martinez-Alonso, Jim Hannigan, David Edwards (ACOM/NCAR); Wolfgang Stremme, Andrea Cadena, Michel Grutter (UNAM); Kimberly Strong, Victoria Flood (U. Toronto); Xiaoyi Zhao (ECCC); Gonzalo Gonzalez Abad, Caroline Nowlan (SAO-TEMPO)

Section 3.3.5 TotHCHO Inter-Comparison Between TEMPO and In situ profiles from NASA DC-8 from AEROMMA

Authored by: **Eleanor Waxman**, Andrew Rollins Kristen Zuraski, Carrie Womack, (NOAA-CSL) Arthur Mizzi, Chia-Hua Hsu, Rainer Volkamer, Sunil Baidar, Carsten Warneke

Section 3.3.6 TotHCHO Evaluation of TEMPO evaluation using WRF-Chem

Authored by: **R. Bradley Pierce**, Jerrold Acdan, Maggie Bruckner (UW-SSEC) Laura Judd, Scott Janz (NASA), Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO)

Section 3.4.1 TotO3 Inter-Comparison Between TEMPO with Polar-Orbiting Satellite Nadir-Mappers

Authored by: **J. R. Ziemke**, S. M. Frith, N. A. Kramarova (NASA GSFC), Kanghyun Baek (Pusan University)

Section 3.4.2 TEMPO ozone validation via ground-based network measurements

Authored by: **Xiaoyi Zhao**, Vitali Fioletov, Debora Griffin, Chris McLinden, Sum Chi Lee (Environment and Climate Change Canada), Irina Petropavlovskikh, (National Oceanic and Atmospheric Administration), Tom Hanisco, (NASA GSFC), Lukas Valin, Eric Baumann, James Szykman (U.S. EPA). Alexander Cede5, Martin Tiefengraber, Manuel Gebetsberger (LuftBlick)

Section 3.4.3 Validation of TEMPO O3 V03 with OMI (Ozone Monitoring Instrument), OMPS (Ozone Mapping and Profiler Suite), and Pandora Ground-based spectrometers

Authored by: **Jay Herman** and Jianping Mao

Section 3.5.1 Background scene assessment of TEMPO , TropNO2, TotHCHO, TotO3

Authored by: **Laura Judd** (NASA LaRC), Caroline Nowlan, Gonzalo Abad (SAO)

Reviewers

Ronald C. Cohen, PhD - Distinguished Professor, Departments of Chemistry and of Earth and Planetary Science, UC Berkeley, Berkeley, CA

David Flittner, PhD - Senior Scientist, Science Directorate, NASA Langley Research Center Hampton, VA

Michael Newchurch, PhD - Professor, Department of Atmospheric and Earth Science University of Alabama in Huntsville, Huntsville, AL

ACRONYMS AND ABBREVIATIONS

AC/VC	Atmospheric Composition/Virtual Constellation
AEROMMA	Atmospheric Emission and Reactions Observed from Megacities to Marine Areas
AGES+	AEROMMA+CUPiDS, GOTHAAM, EPCAPE, STAQS and others
AGL	Above Ground Level
AMAX-DOAS	Airborne Multi-Axis Differential Optical Absorption Spectroscopy
AMF	Air Mass Factor
ATBD	Algorithm Theoretical Basis Document
Cal/Val	Calibration and Validation
CEOS	Committee on Earth Observation Satellites
CTMs	Chemical Transport Models
CTP	cloud-top pressure
CUPiDs	Coastal Urban Plume Dynamics Study
DS	Direct Sun
DSCOVR	Deep Space Climate Observatory
DU	Dobson units
ECC	Electrochemical Concentration Cell
EPA	Environmental Protection Agency
EOL	End of line
EPIC	Earth Polychromatic Imaging Camera
ESA	European Space Agency
EV	Earth Venture
EV-I	Earth Venture Instrument
FOR	Field Of Regard
FOV	Field Of View
FTIR	Fourier Transform Infrared Radiometer
FWHM	Full Width at Half Maximum
GCAS	GeoCAPE Airborne Simulator
GEO	Geostationary Earth Orbit
GEO-CAPE	Geostationary Coastal and Air Pollution Events
GEOS-CF	GEOS Composition Forecasting
GEOTASO	GEOstationary Trace gas and Aerosol Optimization
GLER	Geometry-dependent Lambertian-equivalent surface reflectivity
GOES-R	Geostationary Operational Environmental Satellite - R Series
GOME	Global Ozone Monitoring Experiment
GOME-2	Global Ozone Monitoring Experiment-2

L0	Level 0
L1	Level 1
L2	Level 2
L3	Level 3
LEO	Low Earth Orbit
LIF	Laser Induced Fluorescence
LISTOS	Long Island Sound Tropospheric Ozone Study
MAX-DOAS	Multi Axis Differential Optical Absorption Spectroscopy
NDACC	Network for the Detection of Atmospheric Composition Change
NIER	National Institute of Environmental Research
OMI	Ozone Monitoring Instrument
OMPS-LP	Ozone Mapping and Profiler Suite Limb-Profile
OMPS-NM	Ozone Mapping and Profiler Suite Nadir-Mapper
PBL	Planetary Boundary Layer
PGN	Pandonia Global Network
PLAR	Post-Launch Acceptance Review
PLRA	Program Level Requirements Appendix
PSPIs	Product Specific Performance Indicators
QA/QC	Quality Assurance/Quality Control
RSIG	Remote Sensing Information Gateway
RSME	Root Mean Square Error
S5P	Sentinel-5 Precursor
SAO	Smithsonian Astrophysical Observatory
SCD	Slant Column Density
SHADOZ	Southern Hemisphere ADditional OZonesondes
STAQS	Synergistic TEMPO Air Quality Science
SZA	Solar Zenith Angle
TEMPO	Tropospheric Emissions: Monitoring of Pollution
TOMS	Total Ozone Mapping Spectrometer
TOT	Total Column
TRACER-AQ	Tracking Aerosol Convection interactions ExpeRiment-Air Quality
TROPESS	TRopospheric Ozone and its Precursors from Earth System Sounding
TROPOMI	TROPOspheric Monitoring Instrument
UV/VIS	Ultraviolet/ Visible
UVN	Ultraviolet/Visible/Near-Infrared
VCD	Vertical Column Density
WRF/CMAQ	Weather Research Forecast/Community Multi-scale Air Quality
VZA	Viewing Zenith Angle

EXECUTIVE SUMMARY - TEMPO Level 2 (V3) Data Validation

This validation report highlights new geostationary observations from the Tropospheric Emissions: Monitoring of Pollution (TEMPO) Instrument and the initial performance related to the monitoring column NO2, HCHO and Ozone (Level 2 Version 3, hereafter V3). TEMPO is an UV/VIS imaging spectrometer, hosted aboard a commercial satellite, designed to collect Earth radiance measurements from geostationary orbit to enable the quantification of spatial and temporal variations of trace gases and aerosols at scales relevant for understanding urban air quality in the troposphere.

Current Validation Status: Provisional

This report demonstrates that TEMPO Level 2 Version 3 (V3) products for total column **ozone (O3), tropospheric nitrogen dioxide (NO2), and formaldehyde (HCHO)** meet the required Product Specific Performance Indicators (PSPIs) for **Provisional Maturity** as outlined in the TEMPO validation plan. The data is declared fit for qualitative determinations, of fit-for-purpose, operational testing, and scientific publication.

Performance Against Mission Requirements

Validation was conducted via a multi-faceted approach, including:

- Comparison to ground-based instruments such as Pandora spectrometers and Fourier-transform infrared sites for total and tropospheric column comparisons.

- Evaluation against airborne measurements (e.g., column from NASA GeoCAPE Airborne Simulator and in-situ from the NASA DC-8 aircraft) to assess slant/vertical column densities and vertical profiles.

- Intercomparison with other satellite observations (e.g., TROPOMI) to assess consistency and identify potential biases.

The analyses, with a synopsis in section 2 and details in section 3, evaluated product performance metrics (bias, precision, and uncertainty) and taken together represent a broad a range of conditions and column densities at various spatially and temporally aggregated resolutions.

The results of these analyses were mapped to Product Specific Performance Indicators (PSPIs) and associated Product Maturity Levels (beta, provisional and final) outlined in the TEMPO Validation Plan (NASA, 2023). All analyses contributed to this document were done under a best-efforts approach leveraging measurement and modeling assets and hundreds of scientist volunteers' hours.

Product	PLRA Precision Requirement	Validation Performance (Observed)	Notable Biases/Issues
NO2 (Total and Tropospheric)[1]	1.0×10^{15} molecules/cm^2	TEMPO successfully resolves diurnal changes in pollutant distributions and achieves precision requirement. A multi-year comparison of totNO2 (Aug 2023-mid-July 2025) shows a high correlation with Pandora sites (0.95) and normalized mean bias of -2.9% (-0.27×10^{15} molecules/cm^2) . Results of tropNO2 comparison between TEMPO and TROPOMI, show a R of 0.99 over pandora locations and NMB of 0.012×10^{15} molecules/cm^2. Aircraft comparisons to TEMPO tropNO2 over several urban cites from the NASA GCAS instrument show R^2 ranging from 0.35-0.75 and median difference of -0.4 X 10^{15} molecules/cm^2 to -1.1 X 10^{15} molecules/cm^2, while comparison for totNO2 derived from the DC-8 show a R^2 0.45 and RME 0.76 x 10^{15} molecules/cm^2.	Root mean squared error (RMSE) between TEMPO and Pandora Total NO2 columns shows variations dependent on the NO2 column density levels. Bias peaks in early morning/late afternoon and in highly polluted urban centers.
HCHO (Total)[2]	1.0×10^{16} molecules/cm^2	Observed precision ($<5\times10^{15}$ molecules/cm^2) well within mission requirements. Airborne in situ data tend to agree on uncertainty within the bounds of statistics discussed for Pandora comparisons.	Tends to bias low (31–41%) in high HCHO columns areas (e.g., FTIR site comparisons).
O3 (Total)[1]	3%	Precision: <2%. Accuracy: ±2%. Excellent agreement with ground networks (Pandora, Dobson and Brewer).	Latitudinal bias: High at low latitudes (<25◦N); Low at mid-to-high latitudes (>45◦N).

[1] At spatial scale of less than or equal to 60 km2 at the center of the Field of Regard and a 1-hour temporal revisit time.
[2] At spatial scale of less than or equal to 60 km2 at the center of the Field of Regard and a 3-hour temporal revisit time

Technical Findings & Path to Full Validation

While V3 represents a significant milestone, several issues were identified and are being addressed to reach "Full Validation" status:

- **Cloud Fraction High Bias:** * **Impact:** Affects NO2 and HCHO precision. High cloud bias propagates to Air Mass Factors (AMFs), causing unrealistic spatial variations in trace gas values.
    - **Potential Causes:** High absolute radiance, low surface reflectivity (GLER), or low solar irradiance.

  - **Mitigation:** Apply a conservative **cloud fraction filter (< 0.2)**.
- **Solar Zenith Angle (SZA) Sensitivity:**
  - **Impact:** Retrieval accuracy degrades during early morning and late afternoon.
  - **Mitigation:** Consult the User Guide for **quality flags** to exclude or weight "suspect" retrievals at high SZAs.
- **Total O3 Latitudinal & SZA Bias:**
  - **Issue:** Despite V3 stray light and dark current improvements, a persistent bias remains—high at low latitudes (<25˚ N) and low at mid-to-high latitudes (>45˚ N).
  - **Correction:** Use the **SZA-dependent correction factor** (Eq. 3.4.2.1). With this correction, median bias is generally <3% for SZA < 60˚.

The sections below summarize broad key findings within the report for each TEMPO product. However, more detailed results can be found in Section 2.4 which contains a summary of each in depth analyses conducted by the TEMPO Ad-hoc Validation Working Group members detailed in Section 3.

Next Steps

To achieve full PLRA-required precision and "Final" maturity, the TEMPO Ad-hoc Validation Working Group will continue long-term seasonal characterization and refine the absolute radiance calibrations to mitigate current cloud and SZA biases.

Future Validation Needs

To maximize the utility of TEMPO geostationary observations, ongoing comprehensive validation is

required. This effort must characterize uncertainties at both the **Project Level Requirements Agreement (PLRA)** level and at the finer resolutions used by the research community. While the algorithm team has addressed several initial discrepancies, others require further analysis and will be integrated into future L2 version releases.

Resolved Issues

The synergy between the validation and algorithm teams led to the immediate resolution of two key **Air Mass Factor (AMF)** discrepancies:

- **Tropopause Identification:** Initial difficulties in reconstructing tropospheric and stratospheric AMFs were resolved by updating the algorithm's tropopause identification criteria.
- **Total AMF Calculations:** To reconcile differences between sub-column (stratospheric/ tropospheric) AMF and total AMF calculations, the algorithm team integrated a **temperature correction factor** into the total AMF formula.

These early successes demonstrate the critical value of real-time collaboration between the validation and algorithm development teams.

Acknowledgment

We acknowledge contributions from the measurement community funded outside of the TEMPO mission, including but not limited to the European Union Copernicus Programme, NASA Tropospheric Chemistry Program, U.S. Environmental Protection Agency Air, Climate and Energy Program, NOAA

National Environmental Satellite, Data, and Information Service, Environment Canada and Climate Change Research Program, the National Institute of Ecology and Climate Change (INECC), and the National Autonomous University of Mexico.

1. Objective

1.1 Specific Objectives

This document presents validation results for the Earth Venture Instrument-1 Tropospheric Emissions: Monitoring of Pollution (TEMPO) baseline mission Level 2 (L2) geophysical data products in fashion aligned with the TEMPO Level 2 Science Data Product Validation Plan (citation). That plan provides guidance for validation efforts necessary to establish the validity of TEMPO data products and assigns maturity levels. In particular, the document described analyses of TEMPO data correlation with reference datasets.

1.2 Scope of this Document

This document presents a compilation of scientific analyses submitted on a voluntary basis focused on the validation and quality assessment of the TEMPO Level 2 baseline trace gas data products which include: Total Column NO2, Tropospheric Column NO2, Tropospheric column HCHO, and Total Column O3 for the first year of the TEMPO mission. The Tropospheric Column O3 and 0-2 km O3 (at selected scenes) remain in development and testing phase at the time of this report and are not covered in this validation report.

Over a dozen research groups with expertise on remote sensing retrievals, measurements, and use of modeling in algorithms provided their independent validation analysis. These analyses include the use of independent correlative measurements from satellites, ground-based networks, synergetic science field campaigns along with indirect assessment and evaluation with chemical transport models to validate TEMPO L2 trace gas geophysical data products over a dynamic range of the products and observational conditions, which include solar zenith angles, cloud properties, surface albedo, and abundance of a given pollutant.

The results of each analysis have been mapped to molecule product specific performance indicators (PSPIs), a data qualitative descriptor developed and promulgated in the TEMPO Validation Plan. PSPIs are used to assess and guide a decision on the data maturity level for each TEMPO L2 data product.

The validation maturity of TEMPO Level 2 products is described by three levels: Beta, Provisional, and Full validation. A general description of the three maturity levels follows:

> **Beta:** based on product quick looks using the initial calibration parameters, the product is minimally validated and may still contain significant errors. Publication of research based on Beta maturity products is not recommended and is highly discouraged.
>
> **Provisional:** product performance has been demonstrated through a large, but still (seasonally or otherwise) limited domain of independent measurements. The analysis is sufficient for limited qualitative determinations of product fitness-for-purpose, and the product is potentially ready for testing by operational users and may be suitable for scientific publication.

Full: product performance has been demonstrated over a large and wide range of representative conditions, with comprehensive documentation of product performance, including known anomalies and their remediation strategies. Products are ready for systematic use and cover the full range of the observed domain; products are suitable for basic and applied scientific use and publication.

1.3 Related Documents

TEMPO public release version 3 (V03) L2 and L3 trace gas and cloud product user guide: https://asdc.larc.nasa.gov/documents/tempo/guide/TEMPO_Level-2-3_trace_gas_clouds_user_guide_V1.2.pdf

Validation Plan: https://earthdata.nasa.gov/s3fs-public/2024-05/SAO-DRD-11_TEMPO%20Science%20Validation_Plan_Baseline_1.pdf

Algorithm Theoretical Basis Documents (ATBD)

https://asdc.larc.nasa.gov/documents/tempo/ATBD_TEMPO_v3_O2_Cloud_Product.pdf

https://asdc.larc.nasa.gov/documents/tempo/ATBD_TEMPO_CH2O.pdf

https://asdc.larc.nasa.gov/documents/tempo/ATBD_TEMPO_NO2.pdf

2. Instrument, Level 2 Science Product Description and associated Product Maturity Level Summary

2.1 TEMPO Instrument

The Tropospheric Emissions: Monitoring of Pollution (TEMPO) Instrument [Zoogman et al., 2017] is a dispersive, imaging spectrometer designed to measure solar back-scatter light in the ultraviolet (UV) and visible (VIS) spectral ranges. TEMPO takes advantage of a commercial geostationary host spacecraft to make the first North American tropospheric trace gas measurements from GEO. The TEMPO instrument draws on several decades of heritage spectrometers (GOME, SCIAMACHY, OMI, TROPOMI, GOME-2, and OMPS; Burrows et al., 1999; Bovensmann et al., 1999; Levelt et al., 2018; Munro et al., 2016; Flynn et al., 2014) operating in low-earth-orbit (LEO), adapting and applying the technology for a geostationary satellite mission designed to monitor air quality over North America. Novel to TEMPO are hourly measurements (or less) during daylight hours at high spatial resolution (2×4.75 km^2 at the center of field of regard). The following is a summary of key characteristics for the TEMPO Earth Venture Instrument (EV-I):

- Wavelength range = UV290 – 490 nm and VIS 540 – 740 nm
- Spectral resolution = ~0.6 nm @Full Width Half Maximum (0.54-0.63 nm)
- Two 2048 (spatial) x 1024 (spectral) pixel CCD detectors
- Step/stare 2-axis scan mechanism, with active image stabilization
- Each mirror step is a ~2.85 s snapshot of all 2K N/S cross-track pixels.
- Standard operations include a ~60 minute E-W scan of FOR in 1226 mirror steps
- Orbit = geostationary (35,786 km), 91.0°W above equator
- Calibration wheel with transmissive diffusers for routine solar calibration
- Instrument Control Electronics (ICE) mounted below spacecraft deck
- Images co-added on board before data downlink at native spatial resolution
- Mass: 137 Kg Power: 138W Volume: 1.4m x 1.1m x 1.2 m

Launched on April 7, 2023 the TEMPO instrument first Earth view measurements occurred on August 2, 2023. Nominal operations started on October 19 following the Post Launch Acceptance Review (PLAR). Prior to nominal operations the TEMPO instrument underwent a series of instrument characterization and analysis (ICA) activities. Level 2 data generated prior to nominal operations might have different characteristics (Field of Regard, varying scan time, period of missing data, etc.). Users are advised to review the TEMPO operational daily logs (https://github.com/Smithsonian/TEMPO-Observations-log/blob/main/daily_log.md) for insight on TEMPO daily operations to gain additional insight into potential impact the Level 2 data.

2.2 Schedule

The TEMPO Mission after launch is divided into four operational phases: Launch and Orbit Transfer, Commissioning, Operations, and End of Life. The Commissioning Phase included Spacecraft Commissioning and TEMPO Commissioning. Spacecraft Commissioning included Bus In-Orbit Test (IOT),

Payload IOT, drift to the operational orbit location, and the start of commercial services. TEMPO Commissioning included Activation, On-Orbit Checkout (OOC), and a Post-Launch Acceptance Review (PLAR). Figure 2-1 from the TEMPO Commissioning-Plan (TEMPO-09-0024-TEMPO-Commissioning-Plan_Baseline) below shows the TEMPO Mission Operational Phases. TEMPO launch occurred on April 7, 2023, the TEMPO commissioning phase occurred from June 07-October 16, 2023, and nominal operations began on October 17, 2023 (https://github.com/Smithsonian/TEMPO-Observations-log/blob/main/TEMPO_Daily_Log_during_Commissioning.md)

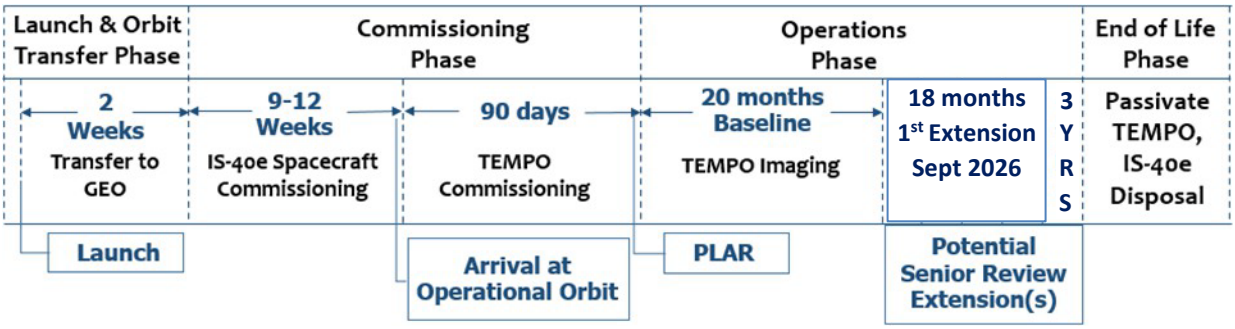


Figure 2-1 Operational Schedule for TEMPO Mission after launch

2.3 Standard Product Requirements

Table 2-1 presents the standard L2 products for ozone, nitrogen dioxide, and formaldehyde, along with product precision and frequency requirements from the TEMPO PLRA version 2 (July 2014) (there might be a more recent version, but the values should be the same). The required precision in Table 2-1 is set for retrievals with 4 native pixels coadded at a spatial scale of ≤ 60 km^2 at the center of FOR. As the spatial resolution of 4 coadded pixels is ~40 km^2, the planned product spatial scales listed Table 2-1 meet or exceed this requirement.

Table 2-1 TEMPO mission Level 2 data product (variable) requirements. Note that the spatial resolution for the precision requirement is with 4 native pixels coadded, not the actual product resolution listed below.

Product Name	Product Horizontal Resolution N/S x E/W @ center of FOR[1]	Product Precision	Temporal Revisit
Total Column O_3	2.0 x 4.75 km^2	3%	1 hour
Total Column NO_2	2.0 x 4.75 km^2	1.0×10^{15} molecules cm^{-2}	1 hour
Tropospheric Column NO_2	2.0 x 4.75 km^2	1.0×10^{15} molecules cm^{-2}	1 hour
Tropospheric column HCHO	2.0 x 4.75 km^2	1.0×10^{16} molecules cm^{-2}	3 hours

[1]Measurement requirements need to be met up to 70° SZA for O_3, NO_2, and HCHO products.

[2]The temporal revisit does not indicate sampling rate as the nominal spatiotemporal sampling is fixed (~1226 E/W mirror steps per hour, with a total of 1226 E/W x 2035 N/S pixels ~2.5M pixels per hour, so that each measurement is a snapshot of ~2.85 seconds) and the nominal product will always be produced hourly. "Temporal revisit" here means the number of measurements (typically hourly) that can be averaged to meet the product precision requirements.

2.4 Product Maturity Level Synopses of Level 2 NO2, HCHO, and O3 data products

This section **provides a synopsis of the validation and quality assessment analyses in Section 3.0 for the Level 2 NO2, HCHO, and O3 data products**. Each summary addresses the relevant Product Specific Performance Indicator addressed by the analysis.

Based on the validation analysis, key algorithm issues identified and addressed (summarized in the Executive Summary). Results indicate PSPIs 01 through 05 and 07 are fulfilled for all the data products, achieving Provisional status. Detailed information is available in the corresponding subsections of Section 3.

2.4.1 TEMPO L2 NO2 Product Maturity Level Synopses

The following section provides a synopsis of the validation and quality assessment analyses for the TEMPO_NO2_Level 2_verision 03 data products from Section 3.2. Table 2-2 list the PSPIs for the NO2 validation levels, with the PSPIs address by each analysis provided at the end of each synopsis.

Table 2-2 TEMPO L2 Total NO2 Product Validation Levels and PSPIs.

Validation Level	Product-Specific Performance Indicators
Beta Validation	NO2-01: Distinguish high NO2 urban areas from nearby rural areas for three select urban-rural scene combinations.
	NO2-02: Assess bias and precision for at least one month of retrievals in comparison to independent correlative measurements to convey an initial characterization to the user community. The assessment should evaluate TEMPO's capability to observe diurnal variations.
	NO2-03: Identify two radiatively homogenous, cloud-clear, low tropospheric NO2 background scenes over a dark surface (e.g., water) and over a bright surface (e.g., snow, desert) under different solar zenith angles and compute point-to-point variability (1-σ) as an empirical estimate for fitting uncertainty. Compare and communicate empirical estimates with those derived from the spectral fitting process.
Provisional Validation	NO2-04: Assess performance metrics (bias/precision/uncertainty) of the tropospheric NO2 product across the CONUS for 1 month period in two seasons, preferably summer and winter, that includes a range of column densities.

	NO2-05: Conduct deep-dive analyses for an episode with relatively poor product performance, identify the root cause and recommend algorithm improvements.
Full Validation	NO2-06: Assess bias, precision, and uncertainty of the tropospheric NO2 product across the CONUS for a wide range of representative conditions over a period of at least one year.
	NO2-07: Assess bias, precision, and uncertainty of the tropospheric NO2 product over areas of interest using data gathered during targeted field campaigns.

Synopsis of Sec 3.2.1 - Comparison: TEMPO L2_NO2 tropospheric column data vs. S5P TROPOMI OFFL L2_NO2 tropospheric column

Scope: TEMPO Field of Regard (FOR); August-December 2023 and May-July 2024

Results: Biases range from $0.75 \pm 28.3 \times 10^{14}$ to $11.3 \pm 34.5 \times 10^{14}$ molecules cm^{-2}, with higher bias in the winter months and lower in the summer months. Correlations are better in the winter months compared to summer and fall, with a peak in November (0.76). TEMPO successfully captures regional variations, recording higher values in urban/industrial areas and lower values in rural/cleaner regions relative to TROPOMI. **Addressed PSPIs:** NO2-02 and NO2-04

Synopsis of Sec 3.2.2 - Comparison: TEMPO L2_NO2 tropospheric column data vs. S5P TROPOMI OFFL L2_NO2 tropospheric column

Scope: Over EPA ozone Non-attainment areas (NAA) and Pandora station locations; August 2023-July 2025

Results: This comparison included an analysis across Pandora sites within the New York-Northern New Jersey-Long Island, NY-NJ-CT and Philadelphia-Wilmington-Atlantic City, PA-NJ-MD-DE NAA to assess as a representation of urban/rural gradients transitions. Pandora station groupings included (1) Cornwall, Madison, New Haven, and Westport, CT; and (2) Philadelphia PA, Bristol PA, New Brunswick NJ, and Bayonne NJ. TEMPO and TROPOMI demonstrated identical rank-ordering of pollution levels across this grouping of urban/rural sites, with TEMPO demonstrating the ability to distinguish varying degrees of pollution across an urban area. TEMPO tended to have higher tropNO2 values compared to TROPOMI in urban environments.

TEMPO and TROPOMI over Pandora sites show a strong spatial correlation of 0.98 with an overall mean bias (normalized mean bias) of 0.012×10^{15} molecules/cm^2 (0.53%). Over Pandora locations, individual site biases between TEMPO and TROPOMI varied with a mean of 1×10^{15} molecules/cm^2 (19%) and a standard deviation of 0.34×10^{15} molecules/cm^2 (14%). For normalized mean bias at NAA regions, there were 37 sites within 15%, 14 sites between 15% and 30%, and no sites above 30%. **Addressed PSPIs:** NO2-01, NO2-02, NO2-04, and NO2-06

Synopsis of Sec 3.2.3 – Comparison: TEMPO L2_NO2 total column data vs. Pandonia Global Network Sites total column NO2 (direct sun)

Scope: Pandora sites across the TEMPO Field of Regard (FOR); August 23-July 25

Results: Box plot comparisons of TEMPO and Pandora medians and interquartile ranges show strong agreement. This demonstrates TEMPO's capacity to distinguish varying pollution levels and capture local urban-to-rural gradients. For the diurnal cycle of totNO2 TEMPO generally remains consistent with Pandora across time of day. However, a stronger low bias at the high-end of the distribution is observed during early morning (06:00–07:00 LST) and late afternoon (15:00–17:00 LST). During the later hours (15:00–17:00 LST), Pandora records a small number of very high totNO2 values that TEMPO does not capture, contributing to the low bias at the high end of the distribution. Despite this, diurnal patterns and correlations remain largely consistent throughout the day. **Addressed PSPIs:** NO2-01 and NO2-02

Synopsis of Sec 3.2.4 - Comparison: TEMPO L2_NO2 tropospheric columns vs. Integrated NO2 in- situ profiles

Scope: NASA DC-8 spirals from NOAA AEROMMA Field Campaign; August 2023

Results: Comparisons were conducted both with and without TEMPO scattering weights and air mass factors (AMF) applied to TEMPO pixels aggregated within the radius of NASA DC8 spirals and vertically integrated in-situ NO2 measurements. The comparisons without TEMPO scattering weights, referred to as "Independent" measurements, were performed to assess the actual errors in the TEMPO tropospheric column retrieval. The TEMPO *a priori* NO2 profile was used above the DC8 in-situ spiral and a constant, boundary layer mean profile was used below the DC8 in-situ spiral. Tropospheric column correlations with and without TEMPO scattering weights and AMF were 0.81 (r^2=0.65) and 0.7 (r^2=0.49), respectively. Tropospheric column Root Mean Error (RME) where 4.97×10^{14} (27.8%) and 9.42×10^{14} (65.7%) molecules cm^{-2}, respectively. Both RMSE values fall within the TEMPO NO2 column precision requirement of 1×10^{15} molecules cm^{-2}. **Addressed PSPI:** NO2-07

Synopsis of Sec 3.2.5 – Comparison: TEMPO TropNO2 vs. NASA GEOCAPE Airborne Spectrometer (GCAS)

Scope: NASA STAQS Field Campaign; August 2023

Results: Both instruments captured major gradients in New York, Chicago, and Toronto. However, TEMPO's dynamic range (5th-95th percentile) for each scan is 38% wider than GCAS over these cities and 67% wider over Los Angeles, indicating higher uncertainty on top of real spatial variance.

A comparison of the population of pixels sampled (N > 5000) shows TEMPO does not appear to have any major systematic biases with respect to GCAS in New York City, Chicago, or Toronto. A similar comparison in Los Angeles (N > 14,000) shows a systematic high bias occurs each of the four flight days. The median difference ranges from -0.4×10^{15} molecules cm^{-2} in Chicago and New York City to 2.0×10^{15} molecules cm^{-2} in Los Angeles. Variance increases (r^2 decreases 0.12 to 0.34 depending on city) between slant column comparisons and vertical column comparisons. The sources of noise will need to be explored within the AMF calculation to identify future pathways for improvement. **Addressed PSPI:** NO2-01 and NO2-07

Synopsis of Sec 3.2.6 – Comparison: TropNO2 Inter-Comparison Between TEMPO and Airborne Multi-Axis DOAS (AMAX-DOAS) measurements

Scope: NOAA CUPiDS Field Campaign: August 2023

Both platforms successfully distinguished NO2 gradients from the urban center of New York metro area to the cleaner rural/coastal areas across Long Island Sound. While TEMPO exhibited larger pixel-to-pixel variability, it accurately captured the same gradient and evolution of the NO2 plumes over the region.

For the limited number of data points and dynamic range of the data assessed (especially at higher solar zenith angles), no major systematic differences were found in TEMPO data with Solar Zenith Angles <70˚. Combined analysis of four Solar Zenith Angle bands (25˚-35˚, 35˚-45˚ degrees, 45˚-55˚ , and 55˚-65˚) showed good agreement (r^2= 0.59, slope of 0.83) between TEMPO and AMAX-DOAS. **Addressed PSPIs:** NO2-01, NO2 -02, and NO2-07

Synopsis of Sec 3.2.7 – Comparison: TEMPO TropNO2 evaluation using WRF-Chem (4 km resolution)

Scope: NASA STAQS and NOAA AEROMMA Field Campaign; August 2023

Results: This indirect validation approach allowed evaluation of TEMPO tropospheric NO2 column retrieval under a wider range of conditions (11 days) as compared to 1-4 days using GCAS direct validation. High resolution (4km) WRF-Chem simulations are used as a transfer standard to conduct comparisons between TEMPO and GCAS TropNO2 measurements. Inferred TEMPO-GCAS biases for Chicago (-8.9x10^{14} +/- 21.x10^{14}), Toronto (-6.7x10^{14} +/- 23.x10^{14}), and New York City (-8.0 x10^{14} +/- 22.x10^{14} molecules cm^{-2}) were all within the TropNO2 product precision requirement. The RMSE between WRF-Chem and TEMPO compared to WRF-Chem and GCAS are very consistent but are 2-3 times larger than the TEMPO NO2 column precision requirement. However, much of this RMSE comes from uncertainties in the WRF-Chem simulations which likely overestimate the true TEMPO RMSE.

TEMPO tropospheric NO2 column retrievals are able to distinguish between urban and neighboring suburban regions in both Chicago, where the TEMPO NO2 column is 50% lower in the Northwestern Suburbs (2.4 x10^{14} molecules cm^{-2}) compared to urban Chicago (4.9 x10^{14} molecules cm^{-2}), and New York City, where the TEMPO NO2 column is 70% lower in Connecticut (1.9e x10^{14} molecules cm^{-2}) compared to urban New York City (6.2 x10^{14} molecules cm^{-2}). Both suburban regions also show less variability compared to their urban counterparts. **Addressed PSPIs:** NO2-01, NO2-02, and NO2-07

Synopsis of Sec 3.2.8 – Comparison: TEMPO GEOS-CF *a priori* profile evaluation vs. DC-8 AGES+ in-situ NO2

Scope: NOAA AEROMMA Field Campaign; August 2023

Campaign-averaged vertical profiles from the GEOS-CF TEMPO *a priori* NO2 and DC-8 AGES+ NO2 observations over urban and rural areas show a high bias in GEOS-CF NO2 predictions within the planetary boundary layer (PBL) and is higher in urban areas. Above the PBL, GEOS-CF exhibited a slight low bias in the free- to upper-tropospheric in both urban and rural areas but generally compared well with the airborne observations. The differences in the ability of GEOS-CF to reproduce observed vertical NO2 profiles in urban/rural regions could lead to differences in TEMPO VCD NO2 biases caused by these shape factor errors resulting in region-dependent AMF errors. **Addressed PSPI:** NO2-07

Synopsis of Sec 3.2.9 – Comparison: TEMPO vs. TROPOMI TotNO2 Vertical Column Density

Scope: TEMPO Field of Regard: May 13- June 16, 2023

TEMPO TotNO2 is 11% lower than TROPOMI on average, with the low bias that is most pronounced over polluted urban areas (e.g., Los Angeles and Chicago). Excluding the highly polluted grid cells (TROPOMI NO2 total columns>$1x10^{16}$ molecules cm^{-2}) improved the R^2 from 0.80 to 0.96. **Addressed PSPIs:** NO2-01, NO2-02

2.4.2 TEMPO L2 HCHO Product Maturity Level Synopses

The following section provides a synopsis of the validation and quality assessment analyses for the TEMPO_HCHO Level 2_verision 03 data products from Section 3.3. Table 2-2 list the PSPIs for the HCHO validation levels, with the PSPIs address by each analysis listed at the end of each synopsis.

Table 2-3 TEMPO L2 Total HCHO Product Validation Levels and PSPIs.

Validation Level	Product-Specific Performance Indicators
Beta Validation	HCHO-01: Distinguish high HCHO concentrations from background concentrations. Given HCHO retrieval noise levels, these qualitative evaluations may use spatial or temporal averaging.
	HCHO-02: Assess bias for at least one month of retrievals, including the diurnal cycle, of comparison to independent correlative measurements to convey an initial characterization to the user community. The assessment should evaluate TEMPO's capability to observe diurnal variations.
	HCHO -03: Identify two radiatively homogenous, cloud-clear, low HCHO scenes over a dark (e.g., water) and over a bright surface (e.g., snow, desert) under different solar zenith angles and compute point-to-point variability (1-σ) as an empirical estimate for fitting uncertainty. Compare and communicate empirical estimates with those derived from the spectral fitting process.
Provisional Validation	HCHO-04: Assess performance metrics (bias/precision/uncertainty) of HCHO product across the CONUS for 1 month period in two seasons, preferably winter and summer, including a range of column densities. This assessment must evaluate the capability of TEMPO to observe diurnal variations of HCHO.
	HCHO-05: Conduct deep-dive analyses for an episode with relatively poor product performance**,** identify the root cause of the discrepancy and recommend algorithm improvements.
Full Validation	HCHO-06: Assess bias, precision, and uncertainty of the HCHO product across the CONUS for a wide range of representative conditions over a period of at least one year.
	HCHO-07: Assess bias, precision, and uncertainty of the HCHO product over areas of interest using data gathered during targeted field campaigns.

Synopsis of Sec 3.3.1 – Comparison: TEMPO L2_TotHCHO column data vs. S%P TROPOI OFFL L2_HCHO column

Scope: Over EPA ozone Non-attainment areas (NAA) and Pandora station locations; August 2023-July 2025

Results: This comparison included an analysis across Pandora sites within New York-Northern New Jersey-Long Island, NY-NJ-CT and Philadelphia-Wilmington-Atlantic City, PA-NJ-MD-DE NAA to assess the ability of TEMPO to distinguish HCHO gradients. Results show for the Pandora stations (Cornwall-CT, Madison-CT, New Haven-CT, and Westport-CT) in the Connecticut NAA box, both TROPOMI and TEMPO demonstrated identical rank-ordering of pollution levels from greatest to least TotHCHO. For the four Pandora stations (Philadelphia PA, Bristol PA, New Brunswick NJ, and Bayonne NJ) within the Philadelphia NAA box, TROPOMI ranked the Bayonne location as highest, with TEMPO placing it as lowest. A comparative analysis of the sensors shows a strong correlation between TEMPO and TROPOMI (r=0.82) when comparing the point locations for Pandoras, but weaker agreement when averaged over the broader NAA regions (r=0.38).

TEMPO and TROPOMI over Pandora sites show a strong spatial correlation of 0.82 with an overall mean bias (normalized mean bias) of -1.5x10^{15} molecules/cm^2 (-15%). Over Pandora locations, individual site biases varied with a mean of -1.3x10^{15} molecules/cm^2 (12%) and a standard deviation of 1.1x10^{15} molecules/cm^2 (10%). At NAA regions, the normalized mean bias was -1.7x10^{15} molecules/cm^2 (-16%) and a standard deviation of 1.1x10^{15} molecules/cm^2 (10%).

TEMPO reproduces some of the spatial variability found in TROPOMI but exhibits large seasonal differences. While agree is closer in the summer, TEMPO retrieves lower values than TROPOMI during winter. It is important to note that TROPOMI has a known high bias (Lambert et al., 2024) at low HCHO and should be viewed here as a correlative measurement rather than validation standard. **Addressed PSPIs:** NO2-01, NO2-02, NO2-04, and NO2-06

Synopsis of Sec 3.3.2 – Comparison: TEMPO L2_HCHO total column data vs. Pandonia Global Network total column HCHO (direct sun)

Scope: Pandora sites across the TEMPO Field of Regard (FoR); August 2023-July 2025

The TEMPO TotHCHO performs reasonably well compared to Pandora as TEMPO captures some of the broad variations across the FoR. However, large biases are seen some sites, particularly in the northeast U.S. Overall, the alignment is notably stronger than the comparison with TROPOMI TotHCHO, which exhibits larger seasonal differences. However, TROPOMI has a known high bias at low TotHCHO (Lambert et al., 2024). While the Pandora HCHO measurement is considered not fully validated it is the official validation product for TEMPO. As a result, TEMPO's good performance against Pandora is weighted more heavily than in comparison with TROPOMI.

Both TEMPO and Pandora evolve similarly through the day, with bother showing lower HCHO in the morning that increases with time-of-day, peaking in the afternoon. The mean bias (normalized mean bias) between TEMPO and Pandora overall was -0.43x10^{15} molecules/cm^2 (-3.9%) and varies by site. The

individual site mean bias (normalized mean bias) had an average of $0.064x10^{15}$ molecules/cm^2 (6.8%) and varies with a mean of $-0.43x10^{15}$ molecules/cm^2 (-3.9%) and a standard deviation of $2.1x10^{15}$ molecules/cm^2 (24%). **Addressed PSPIs:** HCHO-01, HCHO-02, and HCHO-04, and HCHO-06; The results here have also been used to guide analyses for PSPI HCHO-05.

Synopsis of Sec 3.3.3 – Comparison: TEMPO L2_HCHO total column data vs. Pandonia Global Network total column HCHO (direct sun)

Scope: Select Pandora sites across the TEMPO Field of Regard (FoR); August 2023- September 2024

This analysis included the same observational data sets as those used in section 3.3.2 but analyzed different validation metrics. This analysis utilized a subset of 30 Pandora sites down selected using quantitative metrics between direct sun and sky scan measurements, recognizing some of the artificial variations in HCHO that occur in some instruments. Similar to Sec 3.3.2, this analysis found agreement between Pandora and TEMPO total column HCHO with an R^2=0.60 across all sites. A slightly low bias was observed in TEMPO at the highest HCHO columns, with a mean absolute difference of $2.6x10^{15}$ molecules cm^{-2}. Site-by-site, statistics indicate $R^2 > 0.5$ at 19 sites, RMSE < $5x10^{15}$ molecules cm^{-2} for 20 sites and all locations within the precision requirement of $1x10^{16}$ molecules cm^{-2}. The largest diurnal disagreements highlight challenges with individual Pandora sites more than poor performance with TEMPO. **Addressed PSPIs:** HCHO-01, HCHO-02, and HCHO-04, and HCHO-06; The results here have also been used to guide analyses for PSPI HCHO-05.

Synopsis of Sec 3.3.4 – Comparison: T TEMPO L2_HCHO total column vs. Network for the Detection of Atmospheric Composition Change (NDACC) Fourier transform infrared (FTIR)

Scope: FTIR stations located in Mexico City, Mexico; Boulder, Colorado; and Toronto Canada; August 2023 to June 202

The quality of the Level 2 (L2) version 3 (V3) HCHO TEMPO product was evaluated using ground-based solar-absorption FTIR (Fourier Transform InfraRed) measurements with results summarized in Table 2-4.

Table 2-4- Summary of TEMPO versus FTIR median Bias linear Orthogonal Distance Regression for collocated analysis

Station	Bias [TEMPO-FTIR, %]	slope	intercept ($x10^{16}$ / molecules cm^{-2})	r^2-value
Mexico City	-31.0 ± 38.8	0.3	0.90	0.38
Boulder	-41.3 ± 21.3	0.7	-0.06	0.61
Toronto	-37.9 ± 32.2	0.8	-0.11	0.46

Addressed PSPIs: HCHO-01, HCHO-02, and HCHO-04, and HCHO-06

Synopsis of Sec 3.3.5 – Comparison: TEMPO L2_HCHO total columns vs. Integrated aircraft HCHO in-situ profiles

Scope: NASA DC-8 spirals from NOAA AEROMMA Field Campaign; August 2023

Results: Comparisons were conducted both with and without TEMPO scattering weights and air mass factors (AMF) applied to TEMPO pixels aggregated within the radius of NASA DC8 spirals and vertically integrated in-situ HCHO measurements. The comparisons without TEMPO scattering weights, referred to as "Independent" measurements, were performed to assess the actual errors in the TEMPO total column retrieval. The TEMPO *a priori* HCHO profile was used above the DC8 in-situ spiral and a constant, boundary layer mean profile was used below the DC-8 in-situ spiral. Total column correlations with and without TEMPO scattering weights and AMF were 0.63 (r^2=0.40) and 0.57 (r^2=0.32), respectively. Total column Root Mean Error (RME) where $1.31x10^{15}$ (26.1%) and $2.03x10^{15}$ (43.2%) molecules cm^{-2}, respectively. Both RME are within the TEMPO HCHO column precision requirement of $1x10^{16}$ molecules cm^{-2}. **Addressed PSPI:** HCHO-07

Synopsis of Sec 3.3.6 – Comparison: TEMPO TotHCHO evaluation using WRF-Chem (4 km resolution)

Scope: NASA STAQS and NOAA AEROMMA Field Campaign; August 2023

This indirect validation approach allowed evaluation of TEMPO total HCHO column retrieval under a wider range of conditions (11 days) compared to 1-4 days using GCAS direct validation. High resolution (4km) WRF-Chem simulations are used as a transfer standard to conduct indirect comparisons between TEMPO and GCAS TotHCHO measurements. Inferred TEMPO-GCAS biases were Chicago ($0.62x10^{15}$ +/- $5.1x10^{15}$ molecules cm^{-2}), Toronto ($2.83x10^{15}$ +/- $5.1x10^{15}$ molecules cm^{-2}), and New York City ($3.5x10^{15}$ +/- $5.4x10^{15}$ molecules cm^{-2}). These inferred biases are within the TotHCHO product precision requirement of $1x10^{16}$ molecules cm^{-2}. The RMSE between WRF-Chem and TEMPO are lower than the RMSE between WRF-Chem and GCAS and within the TEMPO HCHO column precision requirement of $1x10^{16}$ molecules cm^{-2}. Some of this RMSE comes from uncertainties in the WRF-Chem simulations and so this is likely an overestimate of the true TEMPO RMSE.

The TEMPO HCHO column retrievals can distinguish between high (Eastern VA) and low (Lake Ontario) regions, where the mean HCHO columns are $14x10^{15}$ and $7.0x10^{15}$ molecules cm^{-2}, respectively. The high (Eastern VA) region shows a stronger diurnal variation in the TEMPO HCHO column than the low (Lake Ontario) region. **Addressed PSPIs:** HCHO-02, HCHO-02, and HCHO-07

2.4.3 TEMPO L2 O3 Product Maturity Level Synopses

The following section provides a synopsis on the validation and quality assessment analyses for the TEMPO_O3_Level 2_verision 03 data products from Section 3.4. Table 2-5 list the PSPIs for the O3 validation levels, with the PSPIs addressed by each analysis listed at the end of each synopsis.

Table 2-5 TEMPO L2 Total O3 Product Validation Levels and PSPIs.

Validation Level	Product-Specific Performance Indicators

Beta Validation	O3-01: Distinguish high tropospheric O3 areas resulting from stratospheric intrusion or pollution transport or pollution from nearby normal or low O3 areas for three select high-low O3 scene combinations.
	O3-02: Assess bias and precision for at least one month of retrievals in comparison to independent correlative measurements to convey an initial characterization to the user community. The assessment should evaluate TEMPO's capability to observe diurnal variations.
	O3 -03: Identify two homogenous, cloud-clear, normal tropospheric O3 background scenes over a dark (e.g., water) and over a bright surface (e.g., snow, desert) under different solar zenith angles and compute point-to-point variability (1-σ) as an empirical estimate for retrieval uncertainty. Compare and communicate empirical estimates with those derived from the spectral fitting process.
Provisional Validation	O3-04: Assess performance metrics (bias/precision/uncertainty) of O3 product across the CONUS for 1 month period in two seasons, preferably winter and summer, including a range of column densities. This assessment must evaluate the capability of TEMPO to observe diurnal variations of O3 (profile product only).
	O3-05: Conduct deep-dive analyses for an episode with relatively poor product performance**,** identify the root cause of the discrepancy and recommend algorithm improvements.
Full Validation	O3-06: Assess bias, precision, and uncertainty of the O3 product across the CONUS for a wide range of representative conditions over a period of at least one year.
	O3-07: Assess bias, precision, and uncertainty of the O3 product over areas of interest using data gathered during targeted field campaigns.

Synopsis of Sec 3.4.1 – Comparison: TEMPO TotO3 vs. Polar-Orbiting Satellite Nadir-Mappers (OMPS & OMI) TotO3

Scope: TEMPO Field of Regard (FoR); December 2023 and May 2024

Compared to both OMPS and OMI, the TEMPO V3 total ozone is too high by ~3-5% in low latitudes and too low by several percent in high latitudes. The offsets for TEMPO relative to TROPOMI are smaller than either OMPS or OMI with numbers of +2 to +3 % in low latitudes. TEMPO V3 total ozone relative to OMPS total ozone is too high by 3-5 % for lower solar and viewing zenith angles, latitude, total column ozone, and 331 nm reflectivity and is too high by 3-5% for higher cloud top pressures. **Addressed PSPIs:** O3-02 and O3-04

Synopsis of Sec 3.4.2 – Comparison: TEMPO L2_O3 total column data validation vs. ground-based network measurements and TROPOMI total column ozone

Scope: Brewer, Dobson and Pandora sites across the TEMPO Field of Regard (FoR); August-September 2023 – March-April 2024

TEMPO total column ozone (TCO) v3 data was compared with total ozone measurements from two Dobson sites, four Brewer sites, and 58 Pandora sites and TROPOMI TCO. When compared to the

ground-based networks, TEMPO showed a stronger latitude dependence than TROPOMI. TEMPO TCO data shows reasonable precision and accuracy, with some potential issues that must be addressed in the future. More analysis, such as seasonal (effective temperature) dependency could be examined when more TEMPO observations are available. When compared to ground-based measurements, TEMPO TCO data shows a latitudinal dependency to both the Solar Zenith Angle (SZA) and Viewing Zenith Angle. (VZA) Latitudinal biases, across 6 bins, for SZA range from -1.3% to 3.5% with the highest biases at the southernmost latitudes, and a range of -1.7% to 2.57% for VZA. **Addressed PSPIs:** O3-02 and O3-05

Synopsis of Sec 3.4.3 – Comparison: TEMPO L2_O3 vs. OMI (Ozone Monitoring Instrument), OMPS (Ozone Mapping and Profiler Suite), Earth Polychromatic Imaging Camera (EPIC) and Pandora Ground-based spectrometers total O3

Scope: Comparison over Five Pandora sites; May 2024 (multiple days)

Examples of multi-day time series of TCO data from TEMPO and Pandora at 4 sites (Bayonne, New Jersey, Dearborn, Michigan, Richmond, California, and Toronto-Scarborough, Canada) using 2-day running averages shows that the agreement is well within the 3% goal and most often within 1 to 2%. The TEMPO retrievals over the Southern-most city with a Pandora within its field of regard, Mexico City, suggests that TEMPO may have a latitude dependent stray light residual that has not been completely corrected in V03 as suggested by the parabolic shape centered near solar noon. **Addressed PSPI:** O3-05

2.4.4 TEMPO L2 TotO3, TropNO2, TotHCHO Background Scene Synopsis

The following section provides a synopsis of the validation and quality assessment analyses for the TEMPO_NO2, HCHO, and O3_Level 2_verision 03 data products associated with the analysis conducted specific to PSPI 03 for all trace gases.

Synopsis of Sec 3.5.1 – Comparison: Background scene assessment of TEMPO , TropNO2, TotHCHO, TotO3

Scope: Desierto del Altar (bright desert scene) and Gulf of California and Lake Michigan (water scenes); 30 May 2024

This assessment focused on three geographically distinct, homogenous scenes to minimize natural variability. These scenes did not show significant temporal variance in the L2 products and GOES imagery confirmed 100% cloud-free scenes over the study areas except for the final two hours (16:00-18:00 LT) over the Desierto del Altar. The results show all three products successfully met or exceeded their performance requirements.

TEMPO L2 Product	Observed Variance/Behavior
TropNO2 and TotHCHO	In all scenes the standard deviation is well below the required precision, $1x10^{15}$ molecules cm^{-2} for TropNO2 and $1x10^{16}$ molecules cm^{-2} for TotHCHO, with the typical value of about half the precision requirement in all cases and matching the mean vertical column fitting uncertainty as stated.
TotO3	Variance in all three scenes were under the 3% precision requirement and typically well under 1% except for near the edges of the day (before 06:00 LT and after 18:00 LT in each region).

Addressed PSPIs: NO2-03, HCHO-03, O3-03

3. Validation Results

3.1 Status Summary of Level 1B Radiance and Irradiance

During the update process to Version 3, the TEMPO Level 1b products underwent significant enhancements to improve the radiometric and spectral calibration performance. The Level 0-1 algorithm enhancements led to quality improvements in Level 2 products and enabled the detection of city lights, nightglow, and aurora signals from TEMPO during twilight scans. However, assessments of Version 3 Level 1 data indicate that TEMPO Sun-normalized radiances from the ultraviolet (290–490 nm) and visible (540–740 nm) bands are generally overestimated when compared to radiative transfer calculations and other satellite measurements. The biases are estimated to be on the order of 10%, with possible signal and wavelength dependence. This general overestimation leads to positive biases in Level 2 cloud fraction retrievals, ultimately increasing errors in air mass factor calculations for NO2 and HCHO retrievals. The biases in Sun-normalized radiances are found to have spectral dependence at wavelengths < 325 nm, contributing to latitude-dependent biases in O3 retrievals.
In the visible band, etalon fringe patterns are observed in solar irradiance (IRR) and earthshine radiance (RAD) images, mainly at wavelengths > 650 nm. These residual patterns occur result from imperfect spectral characterization of the radiometric calibration coefficients. The issues described above are primarily caused by discrepancies between the on-ground and in-flight calibration parameters (e.g., solar diffuser transmittance and radiometric calibration coefficients). Additionally, the Sun-normalized radiance biases are partly caused by imperfect stray light correction, which independently affects IRR and RAD and may lead to overcorrection at wavelengths < 305 nm. The on-ground calibration parameters and the stray light correction algorithm are being reassessed, and the results will be incorporated into the next version updates.
For Version 3 Level 1 products, no ad-hoc radiometric adjustments have been made to either the IRR or RAD data. Nor has a correction been applied for the small degradation of the working solar diffuser (~2%/yr at 300 nm and nearly zero in the visible band, with the major changes occurring within the first five months) or the change in telescope and spectrometer responsivity (+/-1%). The Sun-normalized radiance radiometric accuracy varies with latitude (~1%), primarily from variations in IRR, with a slight seasonal cycle. In summary, the radiometric discrepancies noted earlier are fairly constant over time. The diurnal variability of these radiometric discrepancies is under investigation.
The wavelength calibration has detected a gradually increasing shift of up to ~0.08 nm in spectral registration relative to the first-light measurement on August 1, 2023 (defined as the nominal wavelengths), which has been stabilized since around March 2024. This information is included in each Level 1 file. The spectral line shape has proven virtually unchanged since the first-light measurement.

3.2 Level 2 Science Product Validation and Evaluation TEMPO_NO2_L2_V03 - TEMPO NO2 tropospheric and stratospheric columns V03

This report section presents the detailed findings of validation and inter-comparison studies of TEMPO NO2 products with correlative measurements satellites, airborne and ground-based platforms along with chemical transport models conducted by individual research groups.

3.2.1 TropNO_2 Inter-Comparison between TEMPO and TROPOMI

Authored by: Shobha Kondragunta and Zigang Weir (NOAA/NESDIS/STAR)

Introduction

This analysis presents the findings of an inter-comparison study of NO_2 products between the TEMPO (Tropospheric Emissions: Monitoring of Pollution) and TROPOMI (Tropospheric Monitoring Instrument) sensors. This analysis addresses PSPIs NO2-02, NO2-04, and NO2-05.

Since May 13, 2024, TEMPO released its latest version, V03, and back-processed data from August to December 2023. In this report, our inter-comparison focuses on the NO_2 tropospheric column (tropNO_2) to evaluate overall performance. Therefore, a large domain covering 15-68°N, 26-163°W was selected, and all available V03 data were used in this study, including data from August to December 2023 and May 13 to July 29, 2024. Data from TEMPO and TROPOMI were matched within a ±30 minutes time window and aggregated to a 0.1°x0.1° spatial resolution. The inter-comparisons were conducted on a monthly basis. TROPOMI data were screened by qa_value > 0.75, and TEMPO data were screened using three quality flags: main data_quality_flag = 0, solar_zenith_angle < 70, and eff_cloud_fraction < 0.2. In this report, the comparison on the radiances taken from TROPOMI and TEMPO were also conducted in the Baltimore MD area.

Methods

Four methods were used to conduct the inter-comparison:

- **Direct Map Comparison**: Monthly mean tropNO_2 maps were generated for both TEMPO and TROPOMI. Difference maps were created by subtracting TEMPO values from TROPOMI values to visualize spatial discrepancies.
- **Scatter Plots**: Scatter plots were used to perform linear regression analyses, providing correlation coefficients, mean bias, and Mean Squared Deviation Error (MSDE).
- **Quantile-Quantile (QQ) Plots**: QQ plots were constructed to compare the distribution of tropNO_2 values at different pollution levels.
- **Histograms**: Histograms of the matched tropNO_2 values from both sensors were plotted to assess the distribution and create difference histograms to highlight discrepancies.

Results

Monthly mean NO_2 maps revealed regional variations across the CONUS, with high concentrations observed in urban and industrial areas as well as regions affected by biomass burning. The difference maps indicated that TEMPO generally recorded higher tropNO_2 values than TROPOMI in densely populated urban areas, but lower tropNO_2 values than TROPOMI in rural and cleaner regions. Figure 3-1 shows an example from August 2023, with similar results observed in other months.

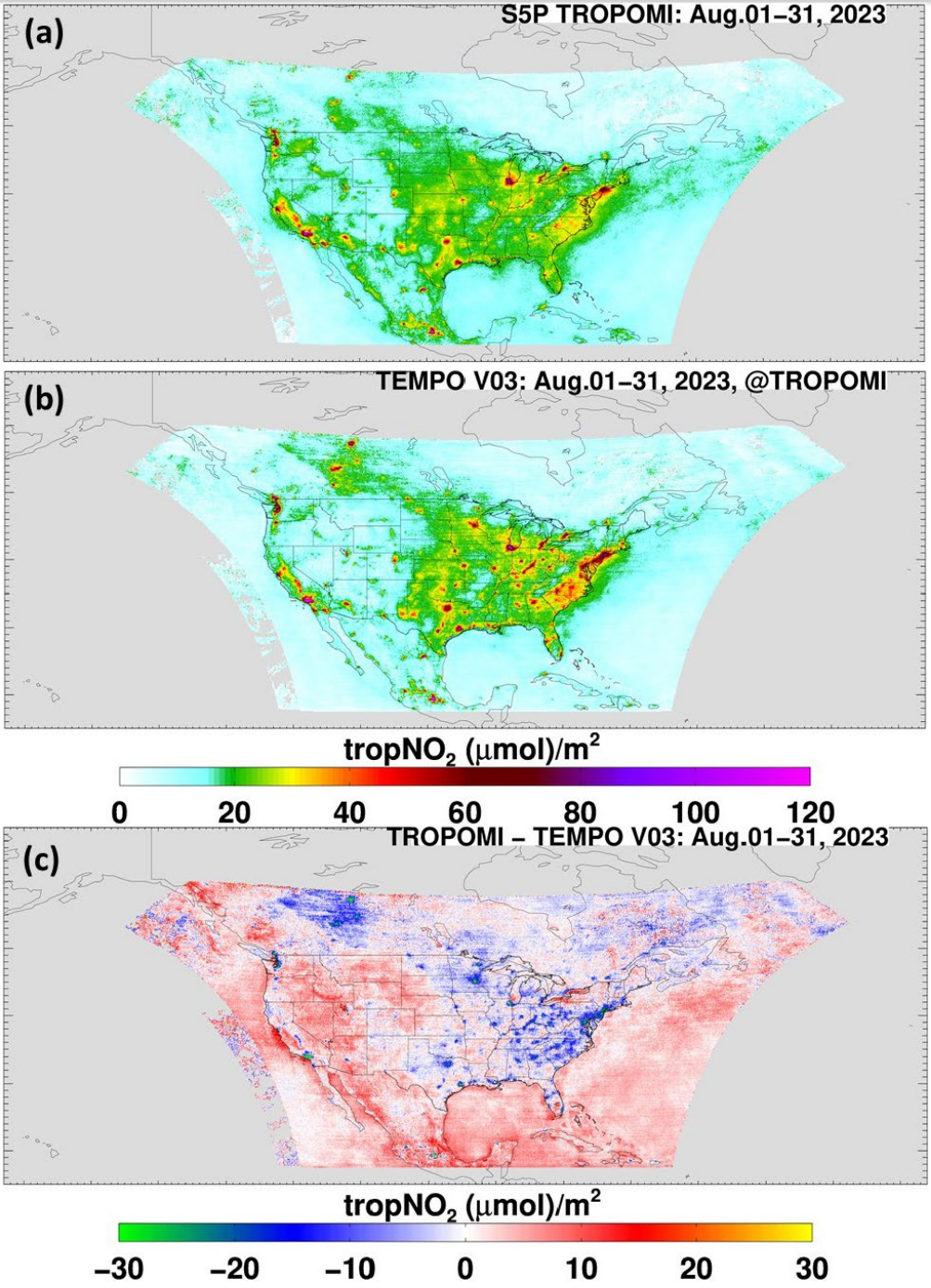


Figure 3-1 Monthly tropNO2 maps for TROPOMI (a), TEMPO (b), and their differences (TEMPO-TROPOMI) (c) in Aug. 2023

Figure 3-2 shows scatter plots for August 2023. The slope of the linear regression was 0.48, and the correlation coefficient was 0.58. Additionally, Figure 3-2 illustrates a mean bias of 1.6 ± 10.5 μmol/m² between TROPOMI and TEMPO for August 2023. The slopes, correlation coefficients, and mean biases for all eight months are listed in Table 3-1. The correlation coefficients between TEMPO and TROPOMI ranged from 0.55 to 0.76 across different months, indicating a high degree of overall agreement. The positive mean biases suggest that TROPOMI consistently reported higher background values than

TEMPO. The correlation between the two sensors was better in the winter months compared to summer and fall, with the highest correlation occurring in December.

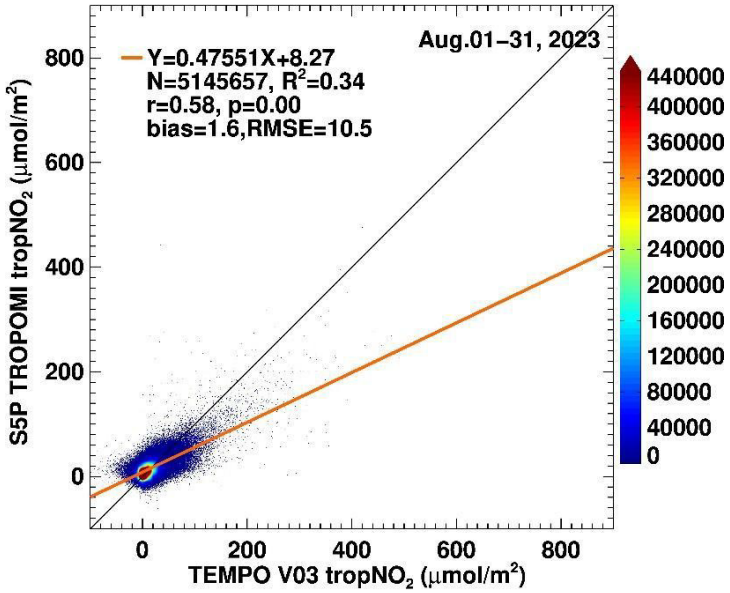


Figure 3-2 Correlations between tropNO2 of TEMPO and TROPOMI in August.

Table 3-1 Correlation Coefficients between TEMPO and TROPOMI tropNO2.

	Aug.2023	Sep.2023	Oct.2023	Nov.2023	Dec.2023	May.2024	Jun.2024	Jul.2024
slope	0.48	0.58	0.67	0.78	0.71	0.50	0.43	0.40
Correl. Coef.	0.58	0.62	0.68	0.76	0.70	0.59	0.55	0.55
Bias (µmol/m²)	1.6±10.5	2.4±10.1	3.2±11.0	4.5±13.8	3.2±15.6	0.3±11.4	0.9±11.9	0.3±11.3

The percentiles at the 1st, 2nd, ..., 99th levels of tropNO$_2$ were taken from the matched grids of TROPOMI and TEMPO for August 2023. These percentiles were plotted as Quantile-to-Quantile plots (QQ-plots), as seen in Figure 3.2.1.3a. The pink dots represent the 10th, 20th, ..., 90th percentiles, marked with thin lines. The QQ plots illustrated that both sensors generally agreed well across different quantiles, though some deviations were noted at the upper extremes of pollution levels. TEMPO tended to measure higher NO$_2$ columns in the upper quantiles compared to TROPOMI and tended to be lower in background values.

Figure 3-3b displays the tropNO2 histograms of the matched TEMPO (cyan curve), TROPOMI (orange curve), and the differences (red curves on the red axis). The histograms demonstrated that the distribution of NO$_2$ values from both sensors was similar, with some differences in the frequency of higher NO$_2$ concentrations. Difference histograms highlighted that TEMPO often measured higher NO$_2$ values than TROPOMI. Similar results were also observed in the QQ plots and histograms (not shown in this report) for the other seven months.

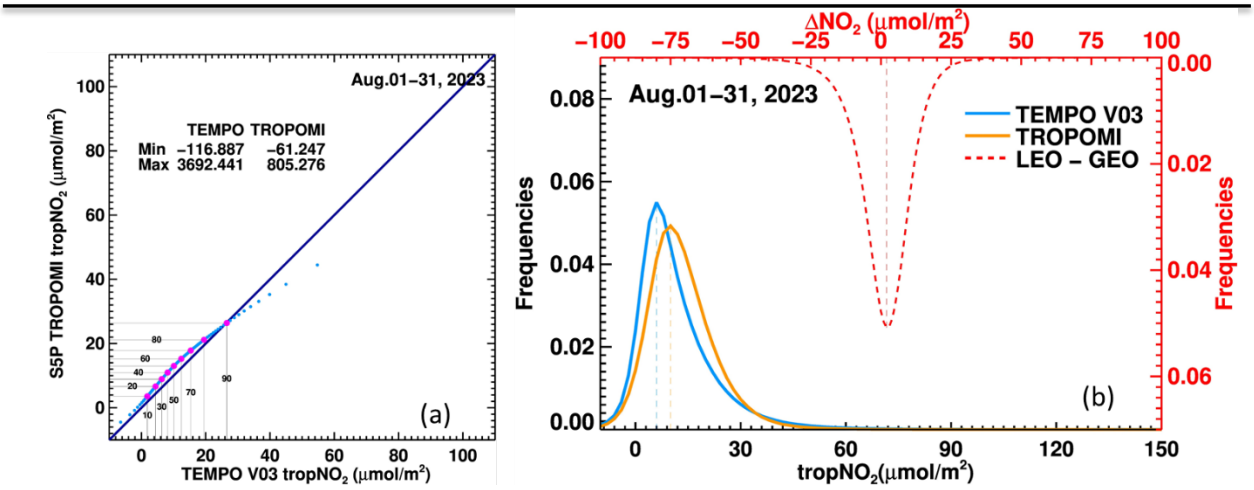


Figure 3-3 tropNO2 Quantitle-Quantitle plots (a) and histograms (b) in August.

To further investigate the causes of the tropNO2 differences between TEMPO and TROPOMI, radiances and irradiances near the NO2 channels were compared, as shown in Figure 3-4. The figure indicates that TEMPO (blue) has a larger radiance (3-4a) in Baltimore, MD, compared to TROPOMI (red), while exhibiting a lower solar irradiance (3-4b). Positive biases in reflectance were observed throughout the NO2 channels, as shown in Figure 3-4c. These differences suggest that further investigation and calibration might be necessary.

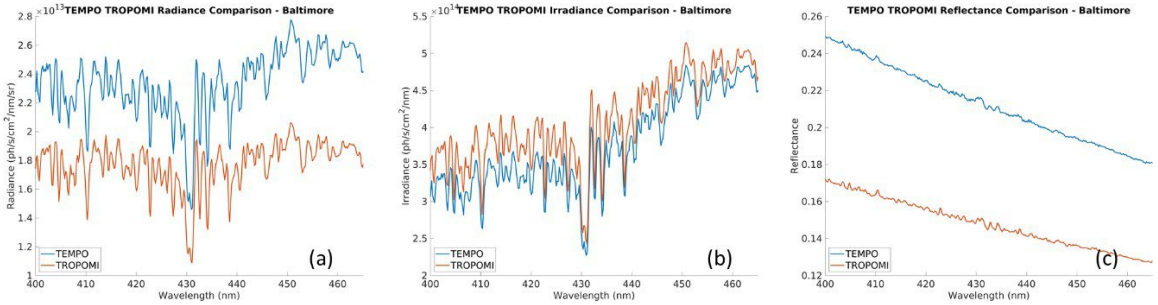


Figure 3-4 Radiance observed by TEMPO and TROPOMI over Baltimore MD.

Discussion

- The monthly mean NO_2 maps highlighted regional variations across the CONUS, with TEMPO generally recording higher tropNO_2 values in urban and industrial areas and lower values in rural and cleaner regions compared to TROPOMI. This could be attributed to differences in sensor calibration, retrieval algorithms, or atmospheric conditions during observations.

- Scatter plot analyses indicated a high degree of correlation between the sensors, with monthly correlation coefficients ranging from 0.55 to 0.76. The results showed that the correlation was stronger in the winter months, with the highest correlation occurring in December.

3.2.2 TropNO2 Inter-Comparison Between TEMPO and TROPOMI

Authored by: Barron Henderson, Lukas Valin, Todd Plessel (EPA-ORD); Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO-TEMPO)

Introduction

This section compares TEMPO to the preexisting TROPOMI NO2 tropospheric product to address PSPIs NO2-01, NO2-02, NO2-04, and NO2-06. The results here have also been used to guide analyses for PSPI NO2-05.

Methodology

Datasets: TEMPO and TROPOMI tropospheric vertical column density nitrogen dioxide (tropNO2) products are coincidently sampled within specific analysis regions. The TEMPO NO2 product that is the focus here is referred to as "version 3" that has been reprocessed to cover the entire period of record (August 2023 to July 2025). TEMPO pixels are only used where the effective cloud fraction is less than 0.15 and QA flags are normal. The TROPOMI product discussed here is the "offline" version that is processed approximately 2-weeks after the measurements are made (S5P_L2__NO2____HiR; C2089270961-GES_DISC). The version of TROPOMI varies slightly over the period of record starting at 2.5.0 in August 2023 and ending at 2.8.0 in July 2025. TROPOMI is filtered using the main data quality flag retaining values greater than or equal to 75 (out of 100). No cloud filter is directly applied to TROPOMI, so all cloud filtering not included in the main quality flag is a result of TEMPO cloud filtering. All data was streamed through the Remote Sensing Information Gateway, which removes negative values. Although this would inherently high bias a dataset, it should high bias all of them similarly.

Sampling: The coincident sampling of TEMPO and TROPOMI has a simple temporal aspect and a more complex spatial component. TEMPO and TROPOMI are compared for all overpasses or scans within the same UTC hour (00:00:00 to 00:59:59Z). Figure 3-5 shows that smaller TEMPO pixels can be intersected with multiple TROPOMI pixels. The TEMPO pixel is ~2 km/pixel in the north-south direction compared to ~5.5km/pixel in the high-resolution TROPOMI along the scanline. The east-west dimension is more comparable with ~4.5 km/pixel for TEMPO and ~3.5 km/pixel for TROPOMI. The actual sizes vary based on viewing geometry within the field of regard. In addition to the difference in sizes, Figure 3-5 highlights that pixels are not aligned resulting in multiple overlaps where valid data is available from both products. Each overlapping intersection is included in the pool of comparisons.

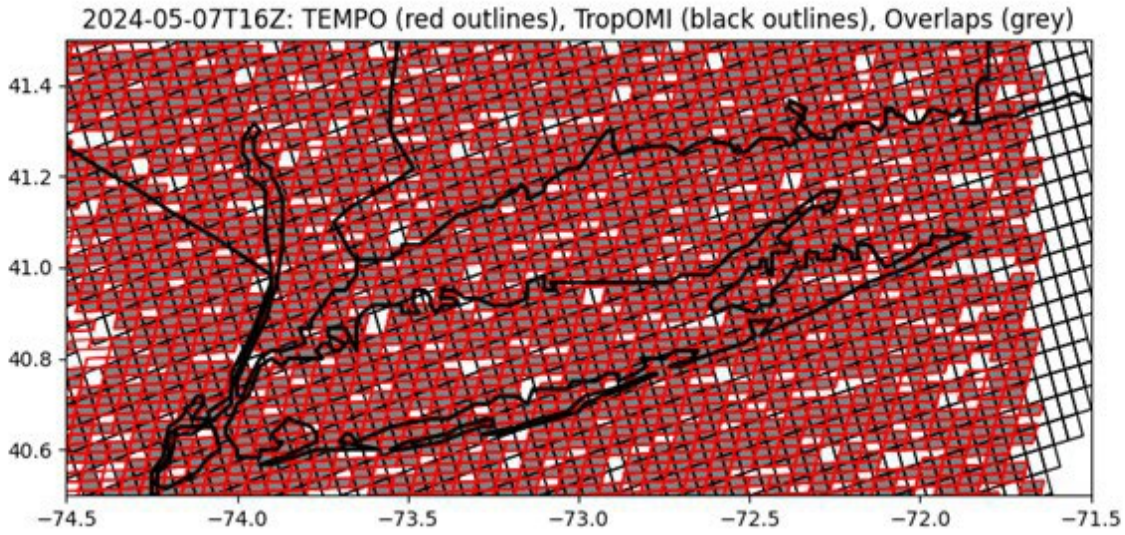


Figure 3-5 Example spatial intersections of TROPOMI and TEMPO over Long Island Sound. The TROPOMI intersections of TEMPO pixel overlap several TROPOMI pixels both due to size and misalignment of the polygons.

Metrics: Bias and precision are key metrics for PSPIs. To address these, we report the mean bias, normalized mean bias, and the Pearson-r correlation metric. These values are assessed at individual analysis regions and across regions. Metrics at an individual site demonstrate how well TEMPO is doing replicating the TROPOMI correlative estimate in specific conditions. Metrics across sites indicate how well TEMPO is able to replicate the variability across locations. It is important to note that replicating TROPOMI is not the same as being correct. Any biases in TROPOMI complicate the interpretation of these results. To highlight this, we perform both linear least-square regressions and orthogonal distance regressions (as implemented in scipy). The difference in fitting parameters highlights the sensitivity to the assumption that the error is attributable only to TEMPO.

Aggregation: The results of intersections are explored at all ozone non-attainment areas (NAA) according to the US EPA National Ambient Air Quality Standards and at Pandora station locations. The NAA regions illustrate TEMPO performance at locations that will be of particular interest in understanding emissions for the purpose of air quality management. The NAA analysis regions are defined by the rectangular envelope that encompasses the nonrectangular region. TROPOMI is also compared to TEMPO over Pandora station locations as context for later comparisons between Pandora and TEMPO. Over Pandora station locations, only TEMPO and TROPOMI pixels within a 0.2-degree box centered on the station are selected. The combined results provide sufficient spatial coverage to address PSPIs. The locations of NAA regions and Pandora stations are shown in Figure 3-6.

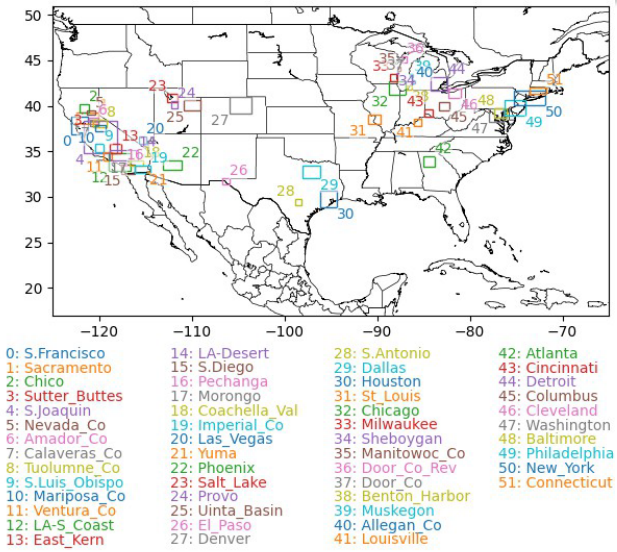


Figure 3-6 Non-Attainment Area (NAA) regions (left) and Pandora locations (right) used in this analysis. The NAA regions are rectangles designed to enclose the NAA polygon. Pandora locations use a 0.2x0.2-degree box around the station.

Results

PSPI NO2-01 requires TEMPO to be able to distinguish urban from rural areas. For PSPI NO2-01, we look at the overall skill of TEMPO tropNO2 NO2 across the US and at individual sites. Figure 3-7 shows TEMPO performance (magnitude, bias, and correlation) at individual regions around the US. Figure 3-8 shows TEMPO and TROPOMI tropNO2 box plot distributions at both Pandora and NAA regions. The NAA regions are spatially much larger than the Pandora regions, which explains why both TEMPO and TROPOMI NO2 variation is larger and more homogenous within the NAA regions. In many cases, Pandora sites are within a NAA region. As a result, the difference between the data at the Pandora locations and the NAA that contains it represents gradients within the NAA. For example, there are four Pandora stations (CornwallCT, MadisonCT, NewHavenCT, and WestportCT) within the Connecticut NAA. Both TROPOMI and TEMPO rank these sites in the same order with the lowest tropNO2 in Cornwall and highest tropNO2 in Westport. Similarly, there are four Pandora stations (PhiladelphiaPA, BristolPA, NewBrunswickNJ, and BayonneNJ) within the Philadelphia NAA. Both TROPOMI and TEMPO rank the four sites in the same order even though TEMPO tends to have higher tropNO2 values compared to TROPOMI. The differences within each NAA highlight that TEMPO is able to go beyond distinguishing urban/rural gradients and can distinguish degrees of pollution within urban areas. Lastly, Figure 3-9 shows a very strong correlation between TEMPO and TROPOMI when comparing across geographical areas. This highlights that TEMPO NO2 is able to capture the spatial variation that is expected from the prior product TROPOMI.

PSPIs NO2-02, NO2-04, and NO2-06 require an assessment of bias and precision for at least one month (NO2-02), two months in different seasons (NO2-04), and a year (NO2-06). This analysis reports several metrics of bias and correlation with respect to TROPOMI for over nearly 2 years of data (Sept 2023 to July 2025). Figure 3-10 provides a temporal overview of the TEMPO and TROPOMI record. This comparison uses interquartile ranges (IQR: 25%-75%) and medians for each month to highlight sub-

seasonal variability. Both TEMPO and TROPOMI capture this sub-seasonal variability with good agreement. The agreement between TEMPO and TROPOMI is shown in Figure 3-11 by month to illustrate the consistency measures of accuracy and correlation. The correlation is broadly consistent with some decreases in summer that are accompanied by TEMPO being higher than TROPOMI from May through August. The dynamic range of observations by location across the entire period can be seen in Figure 3-8. Figure 3-7 and Figure 3-8 show strong spatial correlation between TEMPO and TROPOMI across sites (NAA and Pandora) that is summarized in Figure 3-9. Figure 3-9 shows a remarkable spatial correlation of 0.99 for Pandora sites and 0.92 for NAA . The mean bias (normalized mean bias) overall was 0.012e15 molecules/cm^2 (0.53%) and varied by site. For Pandora locations, individual site biases varied with a mean of 1e15 molecules/cm^2 (19%) and a standard deviation of 0.34e15 molecules/cm^2 (14%). For normalized mean bias at NAA regions, there are 37 sites within 15%, 14 sites between 15% and 30%, and no sites above 30%. For normalized mean bias at Pandora locations, there are 13 within 15%, 17 between 15% and 30%, and 14 above 30%. The correlation also varies by NNA (Pandora) region with 3 (4) above 0.75, 24 (32) between 0.5 and 0.75, and 24 (8) less than 0.5. This highlights that the skill varies by site and users should review the performance at their site. Additional information by location including bias statistics and correlation values is available at https://gaftp.epa.gov/Air/aqmg/bhenders/share/TEMPO/ (available as of 2025-06-26).

In summary, the TEMPO tropNO2 product agrees very well with TROPOMI. The spatial variability has an impressive correlation. The seasonal and sub-seasonal variation are in close alignment and the correlation over months of the year are generally consistent.

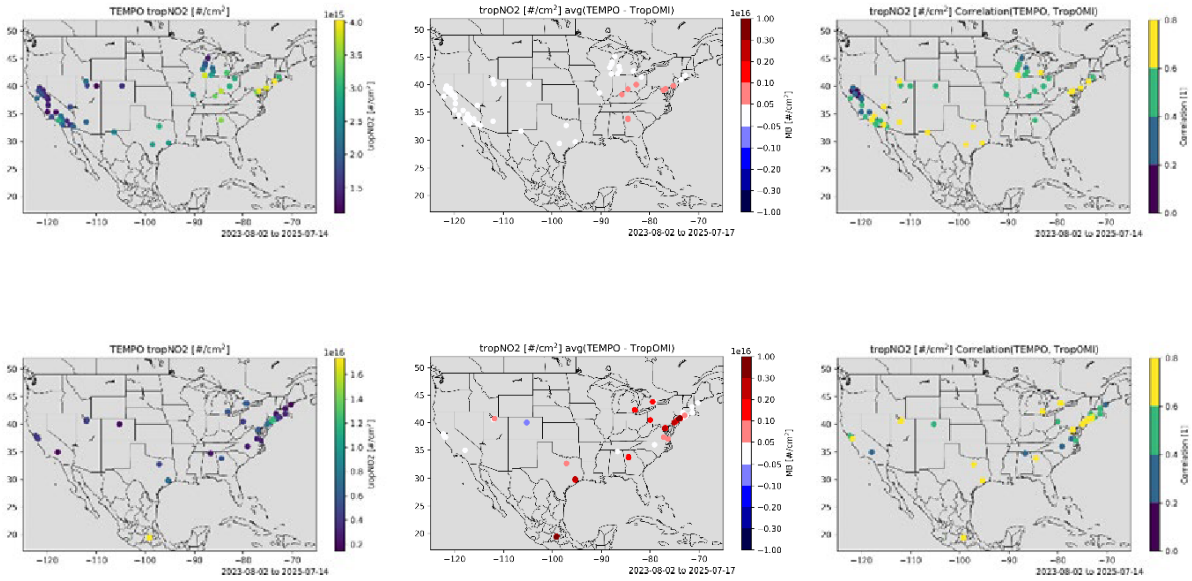


Figure 3-7 TEMPO (left), Mean Bias - TROPOMI (middle) and correlation (right) in tropNO2 at Ozone Non-Attainment Areas (top) and Pandora locations (bottom).

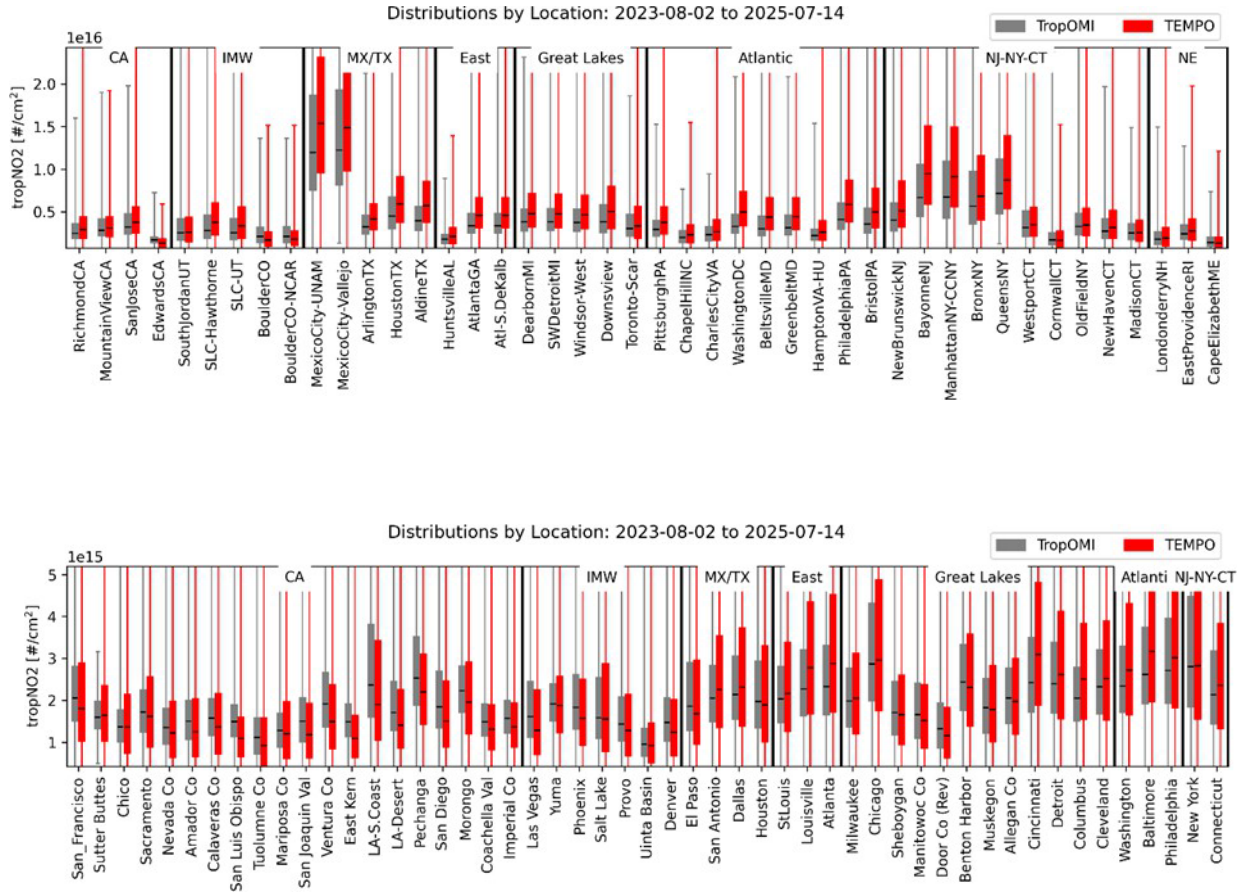


Figure 3-8 TEMPO (red) and TROPOMI (grey) tropNO2 distributions at Pandora (top panel) and Ozone Non-Attainment Areas (bottom panel) locations. The boxes represent the 25%, 50% and 75%, while the whiskers show the minimum and maximum (sometimes of axes). Locations within each panel are sorted from West (left) to East (right).

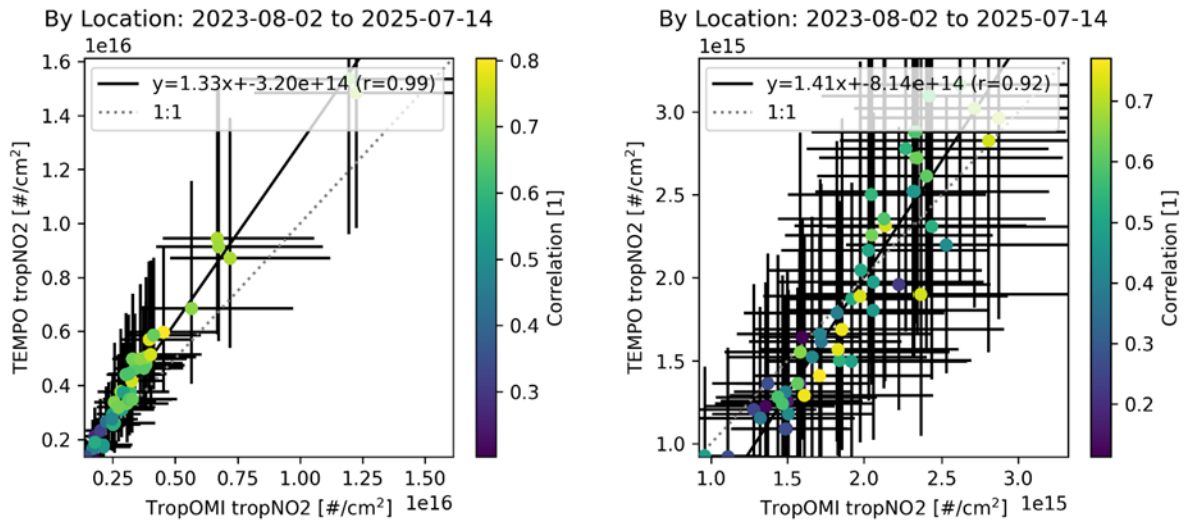


Figure 3-9 TEMPO (y-axis) vs TROPOMI OFFL (x-axis) in tropNO2 for Pandora locations (left) and NAA (right) shows strong correlation. For Pandora sites, TEMPO is higher than TROPOMI particularly at the high-end of the distribution. The legend shows the linear regression (LR).

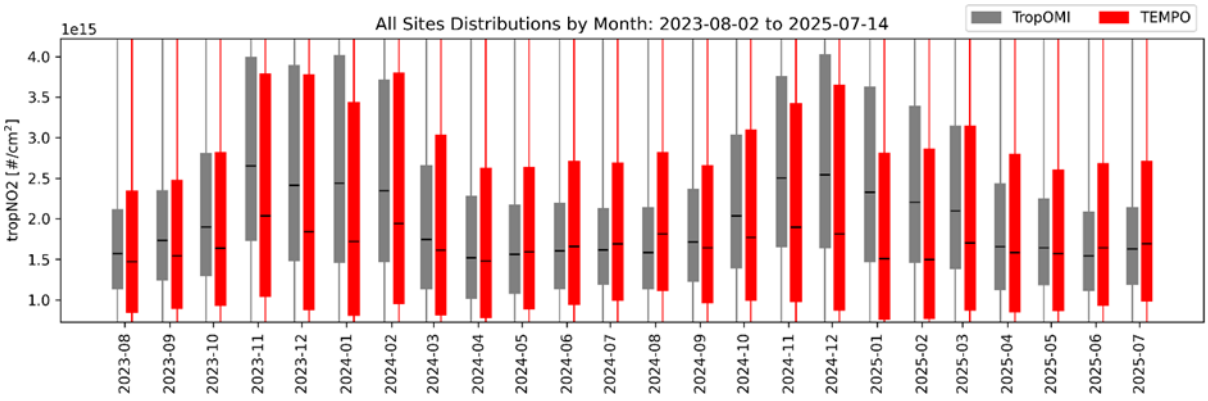


Figure 3-10 TEMPO (red) and TROPOMI OFFL tropNO2 interquartile ranges (line: median, box: 25% to 75%) by month. The whiskers show minima and maxima off scale.

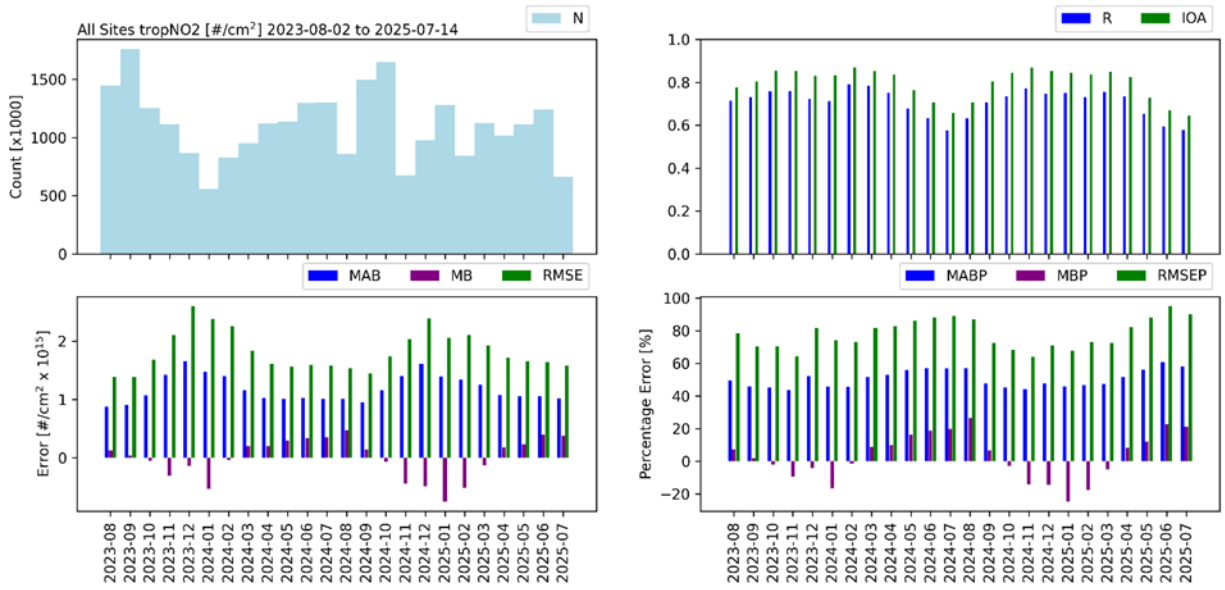


Figure 3-11 Statistics by month from TEMPO and TROPOMI tropNO2 by month. The bottom row shows monthly mean absolute bias (MAB), mean bias (MB), and root mean square error (RMSE) in molecules/cm2 on the left and as a percent of the mean on the right.

3.2.3 TotNO2 Inter-Comparison Between TEMPO and the Pandonia Global Network

Authored by: Barron Henderson, Lukas Valin, Eric Baumann, Dave Williams, Todd Plessel and Jim Szykman (EPA) ; Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO-TEMPO)

Introduction

This section compares TEMPO to the ground-based Pandora NO2 products from the Pandonia Global Network to address validation PSPIs NO2-01, NO2-02, NO2-04, and NO2-06. The results here have also been used to guide analyses for PSPI NO2-05.

Methodology

Datasets: TEMPO and Pandora total vertical column density (VCD) nitrogen dioxide (NO2) products are coincidently sampled at all Pandora sites within the TEMPO field of view. The TEMPO NO2 product that is the focus here is referred to as “version 3” that has data been reprocessed to cover the entire period of record (August 2023 to October 2024). TEMPO pixels are only used where the effective cloud fraction is less than 0.15 and QA flags are normal. The Pandora product discussed here is the “rvns_3.1_8” version that is available through the Pandonia Global Network (PGN) and redistributed through the Remote Sensing Information Gateway. The Pandora NO2 is filtered for medium quality and above based on (Rawat et al., 2025) and any Pandora with valid data within the field of view is included.

Sampling: The coincident sampling of TEMPO and Pandora site has a temporal and spatial filter. Each For each Pandora site (see Figure 3-6), TEMPO and Pandora are sampled within the same UTC hour (e.g. 00:00:00 to 00:59:59Z), but Pandora is likely to have many samples before and after a TEMPO granule. This analysis removes any measurements that are further than 15 minutes from the coincident TEMPO scan. The Pandora data is then paired with any TEMPO pixels that overlap a 0.03-degree buffer around each site. Figure 3-12 shows that this buffer helps to ensure that a Pandora on the edge of two pixels is not artificially favoring one. It also helps to reduce noise in TEMPO retrieval by averaging several pixels. The overall effect of spatial buffering does not meaningfully change the results for NO2, but the same method is used for HCHO where it has a larger impact seemingly associated with noise (see section 3.3.2).

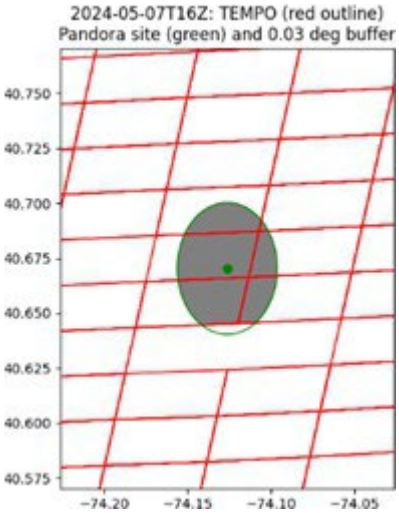


Figure 3-12 Example spatial intersections of TROPOMI and Pandora over Long Island Sound. Several TEMPO pixels overlap a 0.03-degree buffer zone around the Pandora station. Intersections are shown in grey, and the white section of the buffer is where no valid measurement was retrieved.

Metrics: Same as TROPOMI NO2 in Section 3.2.2

Results

Figure 3-13, Figure 3-14, and Figure 3-15 highlight geographic variability of performance relevant to PSPI NO2-01. For PSPI NO2-01, we look at the overall skill of TEMPO totNO2 across the US and at individual sites. Figure 3-13 shows TEMPO performance (magnitude, bias, and correlation) at sites around the US. Figure 3-14 shows the dynamic range of TEMPO and Pandora at each of the locations in Figure 3-13, which highlights that TEMPO is reproducing the appropriate range of values at most sites. The two sites with the largest underestimation are the two in Mexico City, which is the opposite direction in bias as seen by TROPOMI. This difference in sign likely reflects the complexity of this type of analysis from satellite retrievals and potentially spatial heterogeneity within polluted regions (i.e., difference in Pandora view). Lastly, Figure 3-15 shows a very strong correlation (r=0.95) between TEMPO and Pandora

when comparing across geographical areas. This highlights that TEMPO NO2 captures the spatial variation that is expected from the reference product Pandora.

PSPIs NO2-02, NO2-04. And NO2-06 require an assessment of bias and precision for at least one month (NO2-02), two months in different seasons (NO2-04), and a year (NO2-06). This analysis reports several metrics of bias and correlation with respect to Pandora over an entire year (Sept 2023 through Oct 2024). Figure 3-16 highlights sub-seasonal variability with monthly interquartile ranges (IQR: 25%-75%) and medians from both Pandora and TEMPO for all stations. TEMPO and Pandora agree both on the sub-seasonal variability as well as the longer-term seasonal variability. The dynamic range of observations at individual locations can be seen in Figure 3-14. These all show a strong spatial correlation between TEMPO and Pandora across sites that is summarized in Figure 3-15. The mean bias (normalized mean bias) overall is -0.27e15 molecules/cm^2 (-2.9%) and varies by location with a mean of 0.4% and standard deviation of 14%. For the normalized mean bias, 31 sites are within 15%, 12 sites are within 15-30%, and 1 site is above 30%. Similarly, the correlation overall is 0.76 and varies by site (0.58±0.11) with 36 sites between 0.5 and 0.75, and 8 sites less than 0.5. This highlights that the skill varies by site and users should review the performance at their site. Additional information by location including bias statistics and correlation values is available at https://gaftp.epa.gov/Air/aqmg/bhenders/share/TEMPO/ (available as of 2025-06-26).

The direction of the bias is different for Pandora compared to TROPOMI. TEMPO is marginally low-biased compared to Pandora, but high-biased relative to TROPOMI for NO2. The TROPOMI product has a known low-bias (~40% over polluted areas; 13% over clean areas), which likely explains this discrepancy (Lambert et al., 2024).

Figure 3-17 and Figure 3-18 address the diurnal component of NO2-02 by illustrating that TEMPO skill is generally consistent across time of day, though these results may vary by individual location and should be evaluated further. Figure 3-17 illustrates the agreement of diurnal signal by plotting the interquartile range of TEMPO and Pandora totNO2. Figure 3-18 shows that the individual hours show a low bias that is worse at the high-end of the distribution. The 6-7 LST, 15-16 LST and 16-17 LST hours have a stronger low bias than other times of the day. For the later hours (15-17 LST), Pandora has a small number of very large totNO2 values that TEMPO does not capture. In all cases, the least-squares linear regression approach attributes a stronger low-bias to TEMPO than the orthogonal regression. Orthogonal regression assumes uncertainty in both TEMPO and Pandora that may be more appropriate in this case for two reasons. First, both TEMPO and Pandora will have some instrument error. Second, TEMPO and Pandora measure over different areas – attributing all the difference to TEMPO as error could be a mistake. Using the orthogonal regression, the slopes of 6 LST and 7 LST are more similar to other hours and only 6 LST is meaningfully different.

In summary, the TEMPO totNO2 product agrees well with Pandora direct sun measurements. The spatial variability correlates extremely well. The seasonal and sub-seasonal variation is consistent between the two products. The diurnal patterns are consistent on average over all sites, and correlation is largely consistent over the course of the day.

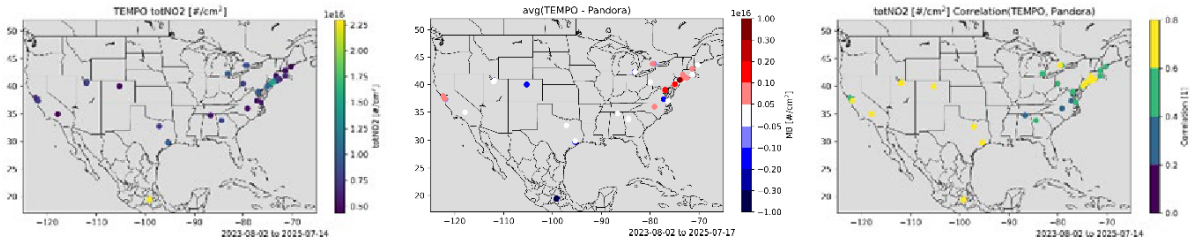


Figure 3-13 TEMPO (left), Mean Bias (middle) and correlation (right) for totNO2 at Pandora locations.

Figure 3.2.3.2: TEMPO (left), Mean Bias (middle) and correlation (right) for totNO2 at Pandora locations.

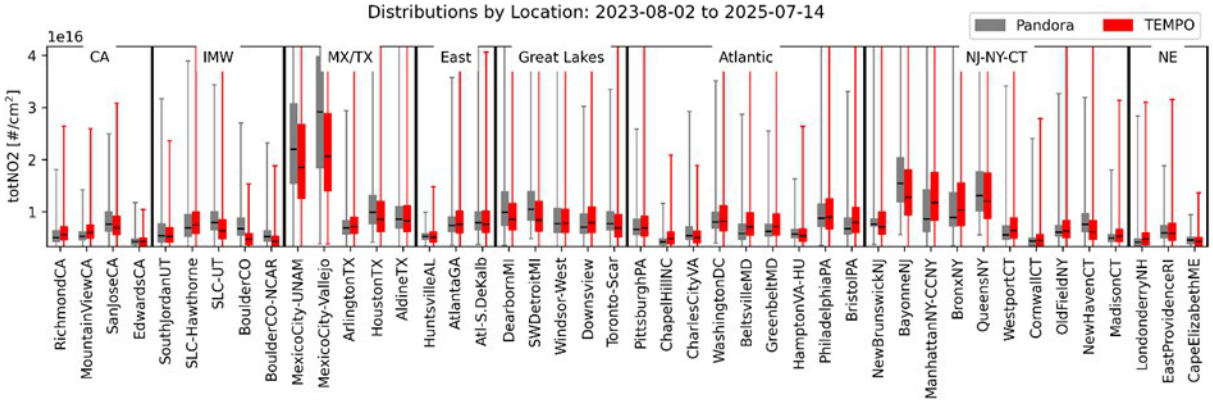


Figure 3-14 TEMPO (red) and Pandora (grey) totNO2 distributions. The boxes represent the 25%, 50% and 75%, while the whiskers show the minimum and maximum. Locations within each panel are sorted into regions: California (CA), Intermountain West (IMW), Mexico/Texas (MX/TX), East, Great Lakes, Atlantic, New Jersey-New York-Connecticut (NJ-NY-CT), and Northeast (NE).

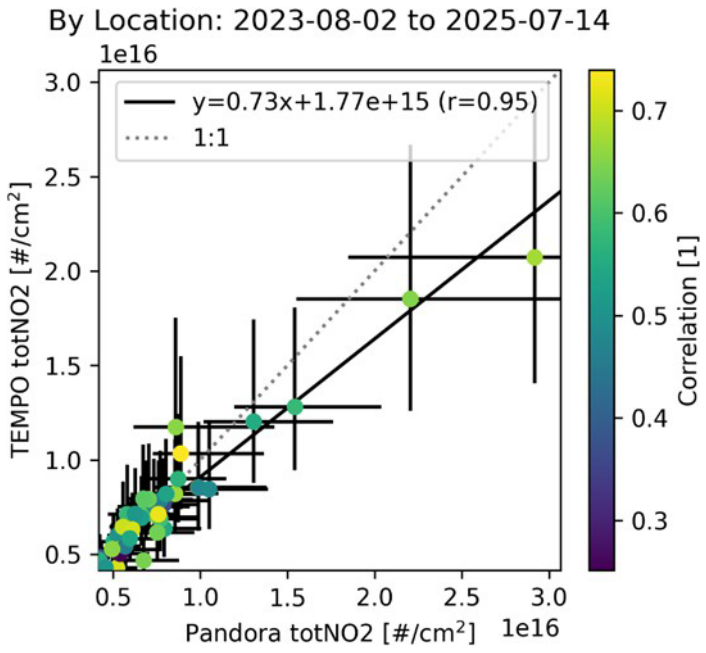


Figure 3-15 TEMPO (y-axis) vs Pandora (x-axis) in totNO2 shows strong correlation with lower TEMPO values for the high-end of the distribution.

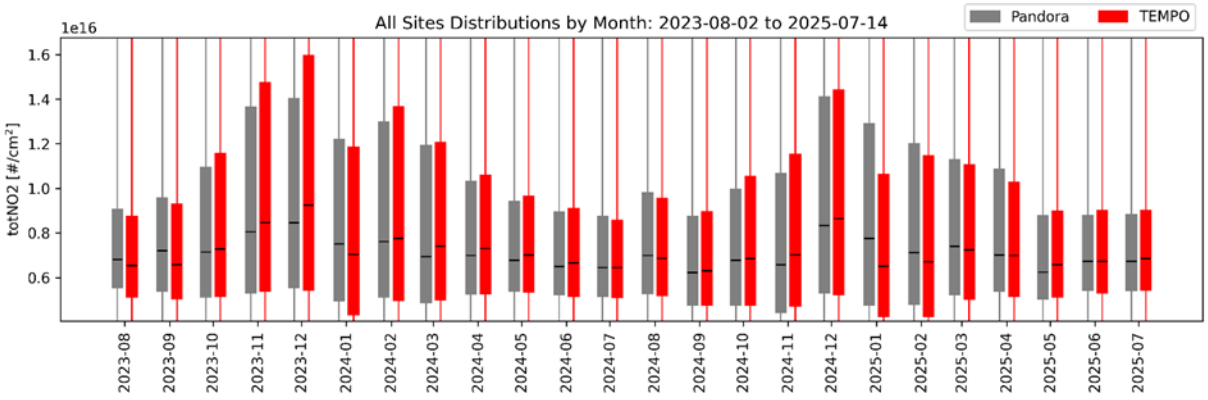


Figure 3-16 TEMPO (red) and Pandora totNO2 interquartile ranges (marker: median, lines: 25% to 75%) by month. Whiskers represent full range and are off the scale.

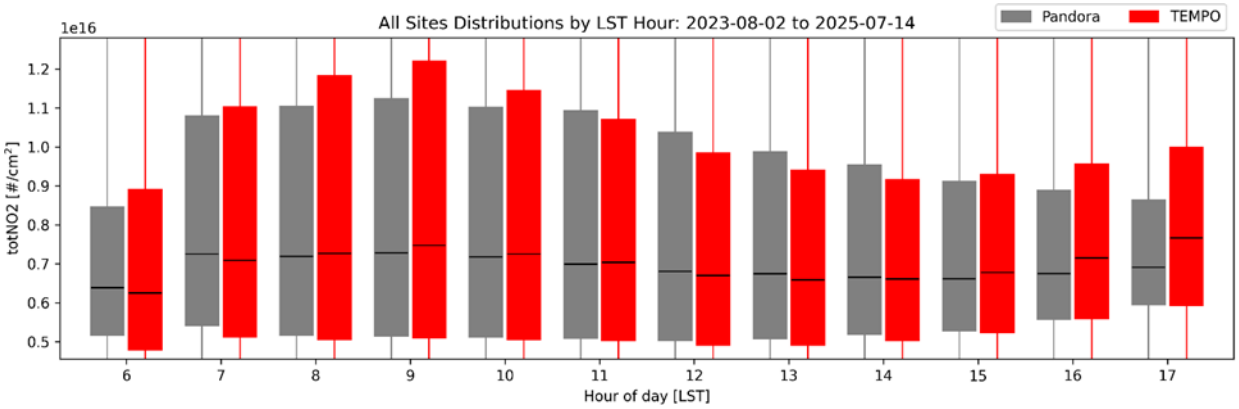


Figure 3-17 TEMPO and Pandora totNO2 interquartile range as a function of local solar time (LST=UTC-LON/15).

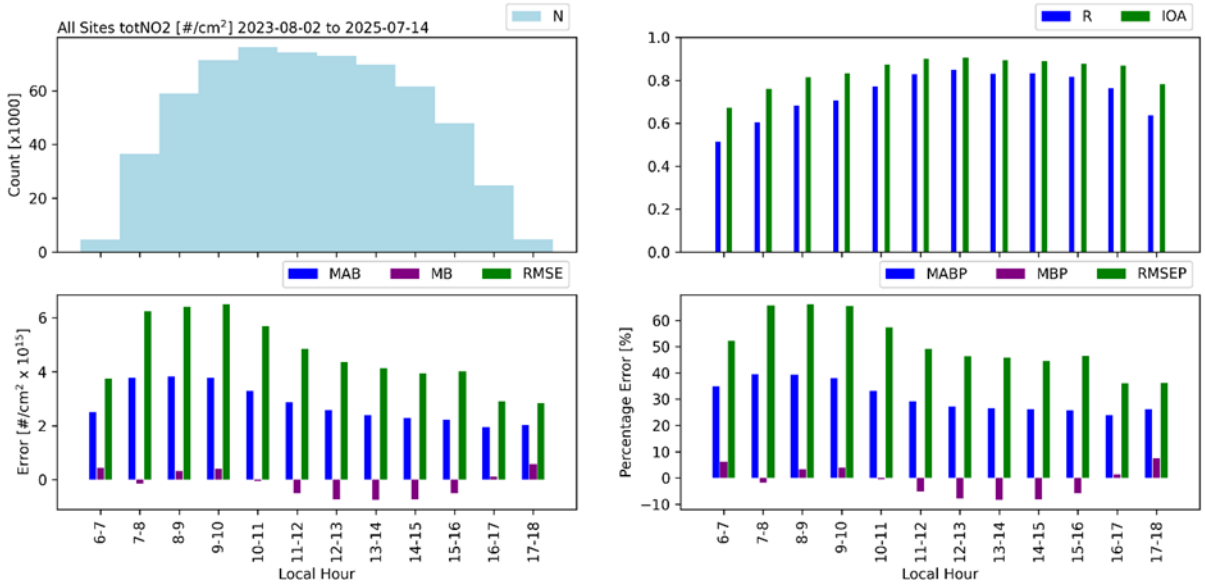


Figure 3-18 Statistics by local hour (LST=UTC-lon/15) for totNO2 from TEMPO (y-axis) and Pandora (x-axis) from 6LST to 17LST across all sites. The top row shows the count (N), correlation (R), and index of agreement (IOA) by month. The bottom row shows monthly mean absolute bias (MAB), mean bias (MB), and root mean square error (RMSE) in molecules/cm2 on the left and as a percent of the mean on the right.

3.2.4 TropNO2 and TotNO2 Inter-Comparison Between TEMPO and In situ profiles from NASA DC-8 from AEROMMA

Authored by: Eleanor Waxman, Andrew Rollins, Kristen Zuraski (NOAA-CSL)

Airborne measurements are largely confined to opportunities linked with air quality field studies rather than systematic collections in time. However, the airborne perspective provides detailed observations of the horizontal and vertical spatial distribution of pollutants to help assess root causes of biases and

uncertainty in TEMPO L2 products and in-situ measurements can enable validation of the slant columns (SCDs) which is unique feature of these measurements.

At the time of writing of this version of the document, all airborne results stem from the summer AGES+ field research effort in August 2023. In summary, within the urban areas of Chicago, Toronto, New York City, and Los Angeles, in situ observations were collected on board the NASA DC-8 from NOAA AEROMMA and airborne remote sensing observations were collected on board two NASA Gulfstreams as part of NASA STAQS. Focused on NYC only, NOAA CUPiDS includes more remote sensing observations. The next three sections offer an in-depth discussion on validation results from each of these airborne efforts.

Dataset Overview: The AEROMMA campaign was a series of flights on the NASA DC-8 during summer 2023 with goals of measuring urban pollution and validating TEMPO products. Flights were made in four urban areas: New York City, Chicago, Toronto/Detroit, and Los Angeles during the month of August 2023. This overlapped with the first month of TEMPO measurements, with first light occurring on August 2, 2023.

For comparison to the TEMPO NO2 product, we use data from the NOAA NOy-LIF instrument on board the DC-8 (modified from Rollins et al., 2020) because it has the best signal-to-noise of the three instruments measuring NO_2 on the DC-8 and it has a unique inlet to minimize interferences in the upper troposphere above about 6-8 km. The instrument measures NO using laser-induced fluorescence by excitation at 215 nm and collection of photons at red-shifted wavelengths. The instrument has been modified since Rollins et al. (2020) to include two additional channels, one to measure NO_x and one to measure NO_y. NO_x measurements are done with the use of a separate inlet that contains a quartz photolysis cell illuminated by two 395 nm LEDs that photolyze NO_2 to NO. This inlet is located outside of the aircraft to ensure that the sample is held at close-to-ambient temperatures to prevent thermal decomposition of NO_y species to NO_2 prior to photolysis. NO_2 is obtained by subtracting the NO measurement from the NO_x measurement. This results in an interference-free measurement of NO_2 in the free troposphere and upper troposphere with an accuracy of 10-20% and 1σ precision of 6.5 ppt at 1 Hz.

Sampling Strategy: During AEROMMA, the NASA DC-8 typically flew between 0.5-12 km above ground level with flight times lasting up to about 8 hours during the daytime hours. A typical flight plan included take off from where the aircraft was based in, Dayton, OH or Palmdale, CA, typically in the late morning hours, a 0.5-2 hour transit at 8-12 km altitude to the urban region of interest, a spiral down to within the boundary layer over the urban area, and several hours of flights in the boundary layer including some spirals between the altitudes of 0.5 to 5 km. At the end of a flight, the aircraft spiraled back up to 8-12 km and had a 0.5-2 hour transit back to the base.

The spirals were designed to provide vertical information for TEMPO validation from in-situ instruments as this provides a vertical profile from about 0.5 km to the free troposphere. These spirals typically encompass 8-25 TEMPO pixels, therefore validation is provided at this multi-pixel scale. We follow an analysis similar to Buscela et al (2008) with the additional comparison including satellite scattering weighting functions applied to the DC-8 data. All results are using TEMPO L2 V03 products.

To identify an in-situ derived profile to compare to TEMPO, we:

- Identify which of those spirals got low enough to enter the planetary boundary layer (using indicators including NO_2, temperature, water vapor, methylvinyl ketone and methacrolein, etc.).

For flights where there is no evidence of a boundary layer, we assume that the aircraft did not enter the boundary layer, and we omit those profiles from our analysis.

- Identify whether there was a TEMPO overpass of the area within 30 minutes of the observed boundary layer by the DC8 with the exception of 25 August 2023 when TEMPO was doing special operations over California, and we restrict this to overpasses within 10 minutes.

Once spirals meeting these criteria are identified, for each spiral each 1 Hz flight measurement is mapped to a TEMPO pixel. Spirals encompassed 8 to 25 pixels. TEMPO pixels are filtered for effective cloud fraction < 0.20 and a main data quality flag of 0. Then the number of times that each TEMPO pixel is sampled is counted. This distribution is used to create weighted averages for TEMPO pressure bins, scattering weights, temperature corrections, air mass factors (AMFs), slant columns (SCDs), and vertical columns (VCDs), etc. for comparison.

Next, the NO_2 measurements from the DC-8 spiral are binned to the average TEMPO pressure bins. Above the maximum aircraft altitude and any gaps in the free troposphere where instrument was being zeroed, we fill in the missing pressure bins with the mean GEOS-CF profile over the pixels. We select the GEOS-CF profile for this substitution as there is generally very good agreement between the in-situ measurements and the GEOS-CF profile above about 3 km when the aircraft is in the free troposphere. Below the aircraft, we substitute in the boundary layer mean from the aircraft measurements. The GEOS-CF temperature profile available in the TEMPO file are also used to apply the temperature correction to the scattering weights as described in the User Guide and Algorithm Theoretical Basis Document for this analysis using the following equation:

$$cc(zz) = 1 - 0.00316[TT(zz) - TT_{\sigma\sigma}] + 3.39 \times 10^{-6}[TT(zz) - TT_{\sigma\sigma}]^2 \qquad (EEEE.\ 3.2.4.1)$$

Where *c(z)* is the vertical profile of the temperature correction, *z* is the pressure layer, *T(z)* is the temperature at layer *z* from the GEOS-CF forecast (provided in the TEMPO files), and $TT_{\sigma\sigma}$ is the temperature of the NO_2 absorption cross section used in the spectral fitting (here, 220 K).

From the airborne data, several datasets are calculated:

- Eq.3.2.4.2 & 3.2.4.3: an in-situ derived total vertical column VCD_{tot_I} and tropospheric vertical column $VVVVVV_{ttttttt\_II}$, calculated through the integration of DC-8 in situ data only,
- Eq. 3.2.4.4: a slant column $SSVVVV_{tttttt}$, derived from the in-situ data and TEMPO scattering weights and the temperature correction (Eq. 1)
- Eq. 3.2.4.5 & 3.2.4.6: a total vertical column VCD_{tot} and tropospheric vertical column $VVVVVV_{ttttttt}$ using the calculated slant column and the mean TEMPO total and tropospheric AMFs.

$$VVVVVV_{ttttt\_II} = \sum_{zz=0}^{ttttt\ ttoo\ aattaattaattheettee} nn_{iiii\ aaiittss}(zz) \times \Delta zz. \qquad (EEEE.\ 3.2.4.2)$$

$$VVVVVV_{tttttttt\_II} = \sum_{zz=0}^{ttttttttttttaassaaee} nn_{iiii\ aaiittss}(zz) \times \Delta zz \qquad (EEEE.\ 3.2.4.3)$$

$$SSVVVV_{ttttt} = \sum_{zz=0}^{tttttt\ ttoo\ aattaattaattheettee} nn_{iiii\ aaiittss}(zz) \times \Delta zz \times SS(zz) \times cc(zz) \qquad (EEEE.\ 3.2.4.4)$$

$$VVVVVV_{ttttt} = \frac{SSVVVV_{ttttt}}{AAAAAA_{ttttt,TTTTTTTTT}} \qquad (EEEE.\ 3.2.4.5)$$

$$VVVVVV_{ttttttt} = \frac{SSVVVV_{ttttttt}}{AAAAA_{ttttttt,TTTTTTTTT}} \qquad (EEEE.\ 3.2.4.6)$$

In the equations above, z represents the pressure layer, $nn_{iiii\ aaiittss}(zz)$ represents the molecular density of NO_2 in each pressure bin z derived from in situ data, Δzz is the thickness of the pressure bin in cm, S(z) are the mean scattering weights reported in the TEMPO file, $cc(zz)$ is the temperature correction factor, $AAAAAA_{tttttt,TTTTTTTTT}$ is the mean total air mass factor averaged from the TEMPO total AMF product, and $AAAAAA_{tttttttt,TTTTTTTTT}$ is the mean tropospheric air mass factor averaged from the TEMPO tropospheric AMF product. The primary difference between equations 3.2.4.2 & 3.2.4.3 and 3.2.4.5 & 3.2.4.6 are that equations 3.2.4.2 & 3.2.4.3 are fully independent from TEMPO data products, while equations 3.2.4.5 & 3.2.4.6 calculate the same final products but include TEMPO data products including the scattering weights and air mass factors.

We do not compare stratospheric VCDs because the DC-8 did not reach the stratosphere and thus there is no added information is available for that part of the atmosphere.

Correlations between TEMPO-derived columns and in-situ-derived columns are fitted with Levenberg-Marquardt least-squares. Y error bars represent the standard deviation of the TEMPO columns that go into the average for each spiral.

Results:

These results address PSPI NO2-07. All AEROMMA spirals have been analyzed using aircraft R0 data available from the publicly available AEROMMA archive (https://csl.noaa.gov/groups/csl7/measurements/2023aeromma/dc8/DataDownload/) and aggregated together for the following results. Figure 3-19 below shows the collection of intercomparisons between the NASA DC-8 derived parameters and TEMPO. Table 3-2 summarizes the statistics for these intercomparisons.

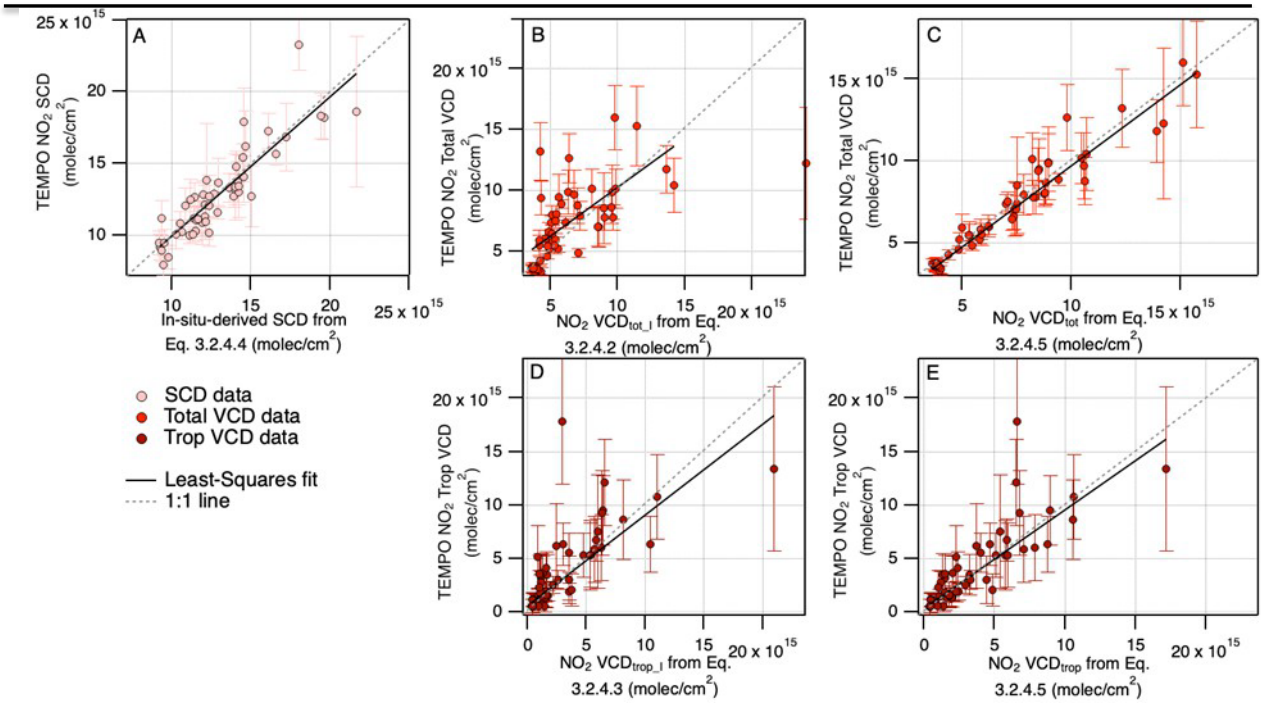


Figure 3-19: Scatter plots of TEMPO vs. DC-8 in situ derived parameters for NO2: (A) TEMPO SCD vs. in-situ derived SCD from DC-8 data using Eq. 3.2.4.4, (B) TEMPO Total VCD vs independent in-situ derived VCD from DC-8 data using Eq. 3.2.4.2, (C) TEMPO Total VCD vs in-situ derived VCD from the DC-8 data and TEMPO priors using Eq. 3.2.4.5, (D) TEMPO Tropospheric VCD vs independent in-situ derived VCD from DC-8 data using Eq. 3.2.4.3, (E) TEMPO Tropospheric VCD vs in situ derived VCD from DC-8 data and TEMPO priors using Eq. 3.2.4.6

Table 3-2 Statistics for TEMPO vs. in-situ derived column comparisons. Note that RME is defined as TEMPO – in-situ, so a positive value indicates that TEMPO is higher while a negative value indicates that TEMPO is lower than the in-situ based columns.

Panel	Compared quantity	Slope (unitless)	Intercept (molec/cm^2)	R^2	RME (molec/cm^2)	RME (%)
A	Slant columns (Eq. 4)	0.97	1.91×10^{14}	0.79	-1.67×10^{14}	-1.3%
B	Total vertical columns (Eq. 2)	0.80	2.27×10^{15}	0.45	7.55×10^{14}	18.2%
C	Total vertical columns (Eq. 5)	0.97	1.46×10^{14}	0.93	-8.88×10^{13}	-1.3%
D	Tropospheric columns (Eq. 3)	0.72	1.89×10^{15}	0.49	9.42×10^{14}	65.7%
E	Tropospheric columns (Eq. 6)	0.90	9.03×10^{14}	0.65	4.97×10^{14}	27.8%

Overall, the correlation between the TEMPO and in-situ SCDs and VCDs is very strong, with R^2 values ranging from 0.45 to 0.93. The mean TEMPO columns on average are slightly higher than the in-situ-derived columns as determined from the RME values, ranging from 1% lower to 65% higher (though we note that no outliers were omitted from the RME analysis so this may be strongly influenced by one or

two points). The strongest agreement is found with the comparison of the total VCDs in Panel C where the slope is near unity and where TEMPO is on average 1.3% lower than the DC-8, which is consistent with the SCD comparison being 1.3% lower. The tropospheric VCD comparisons (panels D and E) had a greater bias of 66% and 29%, respectively. The least correlated datasets are those independent of TEMPO priors (Figure 3-19 B & D) which implies that better agreement in panels C and E is partly due to the assumed vertical sensitivity for TEMPO through their scattering weight calculation. Exploring some of the scatter in the independent vs. dependent calculated profiles and looking at assumptions in the retrieval like clouds, surface reflectivity, etc. can help in understanding whether the differences are driven by the limitations in satellite sensitivity to near surface pollution or errors introduced through assumptions in the derivation of the scattering weights from TEMPO.

Other limitations to this analysis include the flight limitations of the DC-8. In particular, the aircraft was unable to measure within the lowest levels of the boundary layer due to flight restrictions by air traffic control. The highest NO_2 mixing ratios measured by the in-situ instrument were in the boundary layer, so it is important to sample as low into the boundary layer as possible for spirals designed for satellite validation. Several spirals in the current analysis were thrown out because it was not clear that the aircraft was able to get into the boundary layer. In future campaigns, we recommend trying to get as low into the boundary as possible to maximize these important measurements and reduce in-situ-based uncertainty in the boundary layer. Future validation work will include substituting in high-resolution WRF-Chem models for the ~500 m below the aircraft as a sensitivity study on how the results of this analysis change based on assumptions in gap filling.

3.2.5 TropNO2 Inter-Comparison Between TEMPO and NASA GEOCAPE Airborne Spectrometer (GCAS) during STAQS

Authored by: Laura Judd, Scott Janz (NASA); Caroline Nowlan, Xiong Liu, Kelly Chance (SAO-TEMPO)

Dataset Overview: The GEOCAPE Airborne Spectrometer (GCAS) is a UV-VIS instrument capable of mapping spectra across a FOV of 45° nadir of its airborne platform. Because of this capability, GCAS is often referred to as the airborne simulator for TEMPO (and other UV-VIS satellite instruments). Spectra collected in the 425-460 nm window are used to retrieve NO_2 slant columns over areas flown at a resolution of proximately 250 m x 560 m during the STAQS field campaign on the NASA JSC G-V and LaRC G-III. Slant columns are converted to vertical columns through the calculation of the air mass factor as described in Judd et al. (2020) with the exception of using the a priori profile from the GEOS-CF (same prior profile used in TEMPO). GCAS observations have been validated with Pandora spectrometers during previous airborne campaigns, finding close to 1:1 agreement between instruments with uncertainties within ± 25 % and well within than the reported accuracy of Pandora of 0.1 DU (Judd et al., 2020, Nawaz et al., 2024). During STAQS, comparisons from 120 coincidences over 15 Pandora locations measuring direct sun columns show an r^2=0.90 and a reduced major axis linear slope of 1.00 and small positive offset of $1x10^{15}$ molecules cm^{-2}. The median percent difference is 12% with an interquartile ranging from 0-25%. The source of this small high bias is likely due to assumptions in the reference spectrum and will be considered in the TEMPO bias analysis herein. More direct-sun Pandora data is expected within the STAQS GCAS domain as newer instruments within PGN as they are calibrated and processing vertical column data (e.g., ChicagoIL; WhittierCA). Results will be updated as they become available.

In this document, all results with GCAS are evaluating V3 of the TEMPO Tropospheric Vertical Column.

Sampling Strategy: Over 11 days in August 2023, GCAS flew as TEMPO scanned across its field of regard. Cities mapped during these flights include New York City, Chicago, Toronto, and Los Angeles. During flights, GCAS would 'lawnmower' across the city of interest at FL280 (~8.5 km) over the course of 2-2.5 hours (8-10 flight lines of about 135 km in length and spaced 6.3 km apart, collectively referred to as a 'raster'). GCAS flight patterns were repeated up to 3x in a flight day spanning from early morning through late afternoon.

TEMPO data in the results below have a main_data_quality_flag equal to 0 and have eff_cloud_fractions less than 0.15, as is recommended by the TEMPO NO_2 and HCHO User Guide. Unless stated otherwise, to quantitatively match GCAS data with TEMPO, we consider data temporally coincident if measured within ± 30 minutes from the TEMPO observation (as was found to be appropriate for TROPOMI in Judd et al., 2020). Further temporal constraints can be explored in the future. GCAS must also mostly map the TEMPO pixel area with cloud-free data (herein filtered at 75% mapped).

Metrics: Quantitative metrics include calculation of relative and absolute median and percentile differences between GCAS and TEMPO (TEMPO minus GCAS). Regressions are all reduced-major axis recognizing that both measurements have errors stemming from uncertainties in the spectral fit for NO_2 as well as assumptions taken in the air mass factor calculations. Coefficients of determination (r^2) are also provided whenever there a regression is used. Some analysis is also qualitative in nature.

Results: Figure 3-20 illustrates how both TEMPO and GCAS can distinguish gradients spanning from high NO_2 over the emission sources in each city to the cleaner scenes around the edges of the cities—satisfying **PSPI NO2-01**. In these cases, TEMPO captures the same spatial peaks and valleys of the tropospheric NO_2 in the urban domains as GCAS. Note that this is a case where the difference in time between observations may be as large as 1.5 hours so that may explain some of the small differences. Visually, there is also more noise in the NO_2 column from TEMPO than from GCAS.

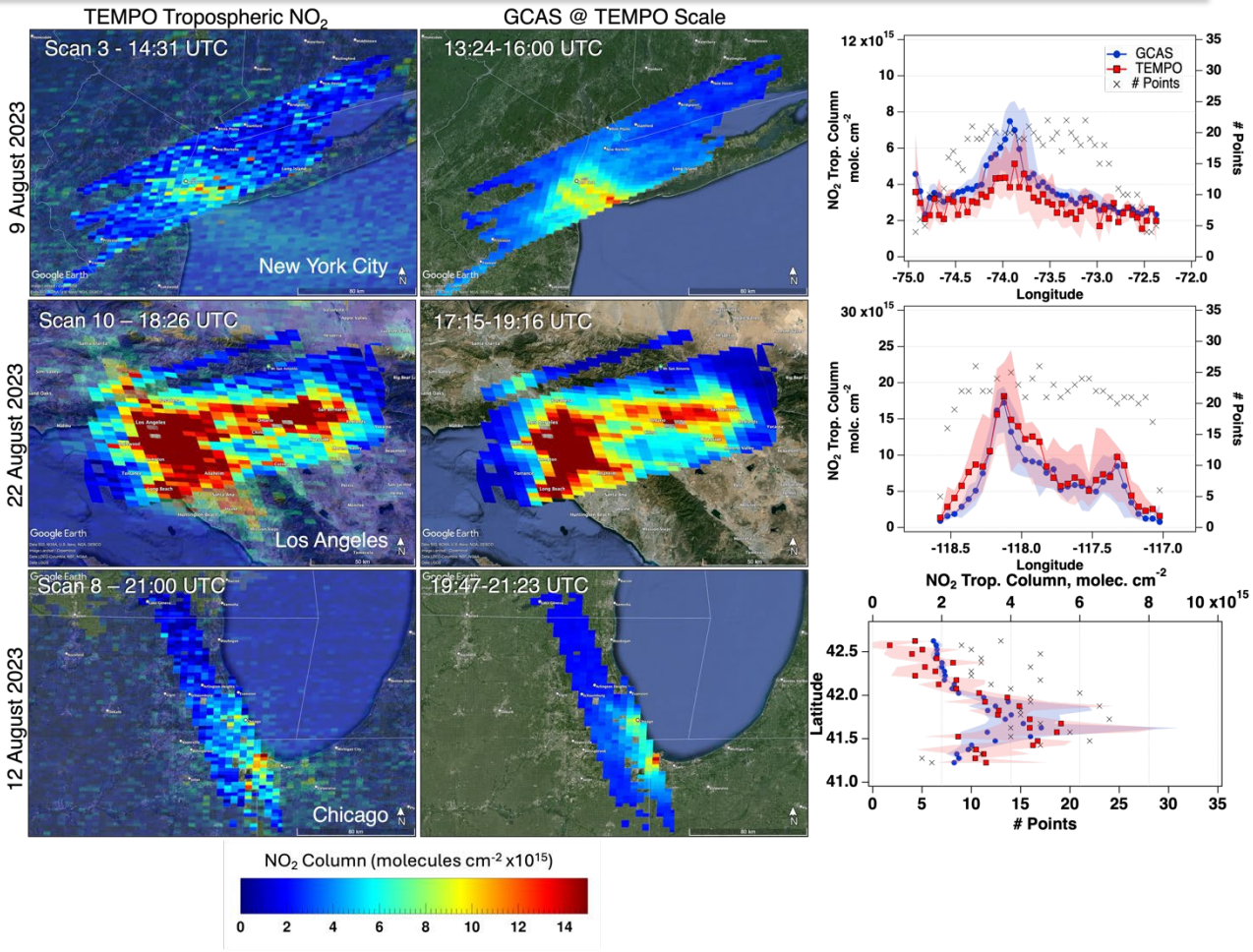


Figure 3-20: TEMPO tropospheric NO_2 columns vs. the most coincident GCAS raster on the noted day and time. The top row shows August 9th, 2023, in New York City in the morning. The middle row shows Los Angeles data around midday on August 22nd, 2023. The bottom row shows Chicago in the mid-to-late afternoon on August 12th, 2023. The left column shows TEMPO tropospheric NO_2 data during the scan closest to the midpoint time of the GCAS raster, where the opaque data was at least 75% mapped during the coincident raster map collected from GCAS. The middle column shows the GCAS data mapped to the TEMPO pixel sizes during the coincident raster. The right plot shows the median and the interquartile of TEMPO (red) and GCAS (blue) data binned in 0.05-degree increments along the predominant latitude/longitude axis of the raster (longitude for NYC and LA; latitude for Chicago). The number of points in each bin is shown by the black x's.

Figure 3-21 illustrates a more quantitative way to compare the dynamic range in NO_2 observed by data coincidences between GCAS and TEMPO. This figure collates data by individual TEMPO scan and subsetting coincident GCAS data collected within ± 30 minutes. Each point represents the dynamic range of that scan from the 5th-95th percentile observed by each instrument during that TEMPO overpass. Scans with less than 20 pixels were filtered from this analysis. The average number of pixels per scan was 209 with a maximum of 341. The dynamic range was quite different for Los Angeles (right) compared to the other three cities (left). These results show that the dynamic range of data from both platforms is highly correlated (r^2=0.64, 0.75 in the eastern cities and Los Angeles, respectively). The dynamic range in the eastern cities is 38% wider in TEMPO than in GCAS but this follows the 1:1 line likely representing

uncertainty in TEMPO. This number is much larger in Los Angeles (67%). The NO_2 tropospheric column comparisons will reveal a noted bias in Los Angeles during flight days that is not apparent in any of the other cities (more below).

This investigation is at the city scale; future analysis should consider looking at gradients across neighborhood scales to ensure analysis at that level for applications like quantifying air pollution disparities and identifying small scale emission plumes are appropriate.

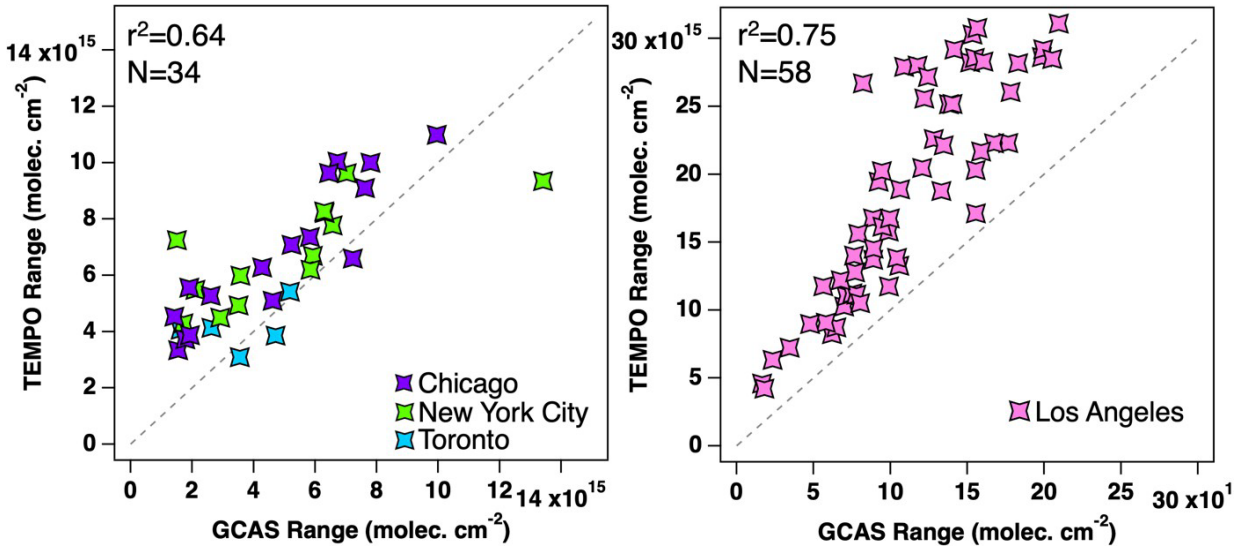


Figure 3-21: Scatter plot showing the dynamic range (defined as the 5th-95th percentile range) of tropospheric NO2 from GCAS and TEMPO for each individual TEMPO overpass colored by region. Each point is colored by city region the data was collected. Chicago, New York City and Toronto are shown on the left with a smaller axes scale and Los Angeles is on the right.

Figure 3-22 shows the collective results for all coincidences following the criteria discussed in the *Sampling Strategy* section above for all flights separated by region (**addressing PSPI NO2-07**). Given the volume of data, it is represented as a density plot for all individual coincidences. Each TEMPO scan is represented by the white dots where the median TEMPO and GCAS value are compared for all pixels in the scan with the bars representing the interquartile range.

New York City, Chicago, and Toronto all have a density of data that surrounds the 1:1 line with slopes ranging from 0.89-1.15. Los Angeles exhibits a high bias in TEMPO data in relation to GCAS. GCAS has a close to 1:1 relationship with coincident Pandora sky-scan data and TROPOMI tropospheric vertical columns at this time in Los Angeles, suggesting that the TEMPO retrieval is the cause for the bias. This high bias does not appear to always be evident though as shown by the TROPOMI/TEMPO results in section 3.2.2 . The specific causes for this high bias will be further explored for PSPI NO2-05.

New York City, Chicago, and Toronto all exhibit a small negative median difference (ranging from -0.4×10^{15} to -1.1×10^{15}) that is within the uncertainty of GCAS and Pandora reference spectrum assumptions, which may be on the order of 1×10^{15} as discussed in the Pandora/GCAS comparisons above.

R^2 ranges from 0.35-0.75 depending on the region. The lowest r^2 is from the Toronto flight, which had the fewest number of cloud-free pixels by far and the lowest dynamic range. New York City and Chicago have a similar number of coincident pixels to each other, but NYC exhibits a larger r^2 likely due to a larger spread

in the density of points above $5x10^{15}$ molecules cm^{-2}. Los Angeles has the largest number of coincident pixels and the widest dynamic range. This is the most polluted area sampled during these flight days and included one day of 10-minute scans from TEMPO (August 23rd), boosting the number of data points. R^2 increases in all cases if comparing the GCAS differential slant column to the tropospheric slant column from TEMPO (vertical_column_troposphere * amf_troposphere)—r^2=0.90 in Los Angeles; 0.68 in Chicago; 0.62 in Toronto; and 0.72 in New York City. For PSPI NO2-05, the added variance in the datasets due to the application of the air mass factor will be explored.

At a high level,

- TEMPO does not appear to have any major systematic biases in New York City, Chicago, or Toronto on the days sampled, but the data in Los Angeles is an example where systematic high bias can occur.
- Any negative offsets are well within the uncertainty for the GCAS reference spectrum and should not be interpreted as an offset bias.
- Variance in the comparison (r^2 decreases) between slant column comparisons and vertical column comparisons. The sources of noise will need to be explored within the AMF calculation to identify future pathways for improvement.

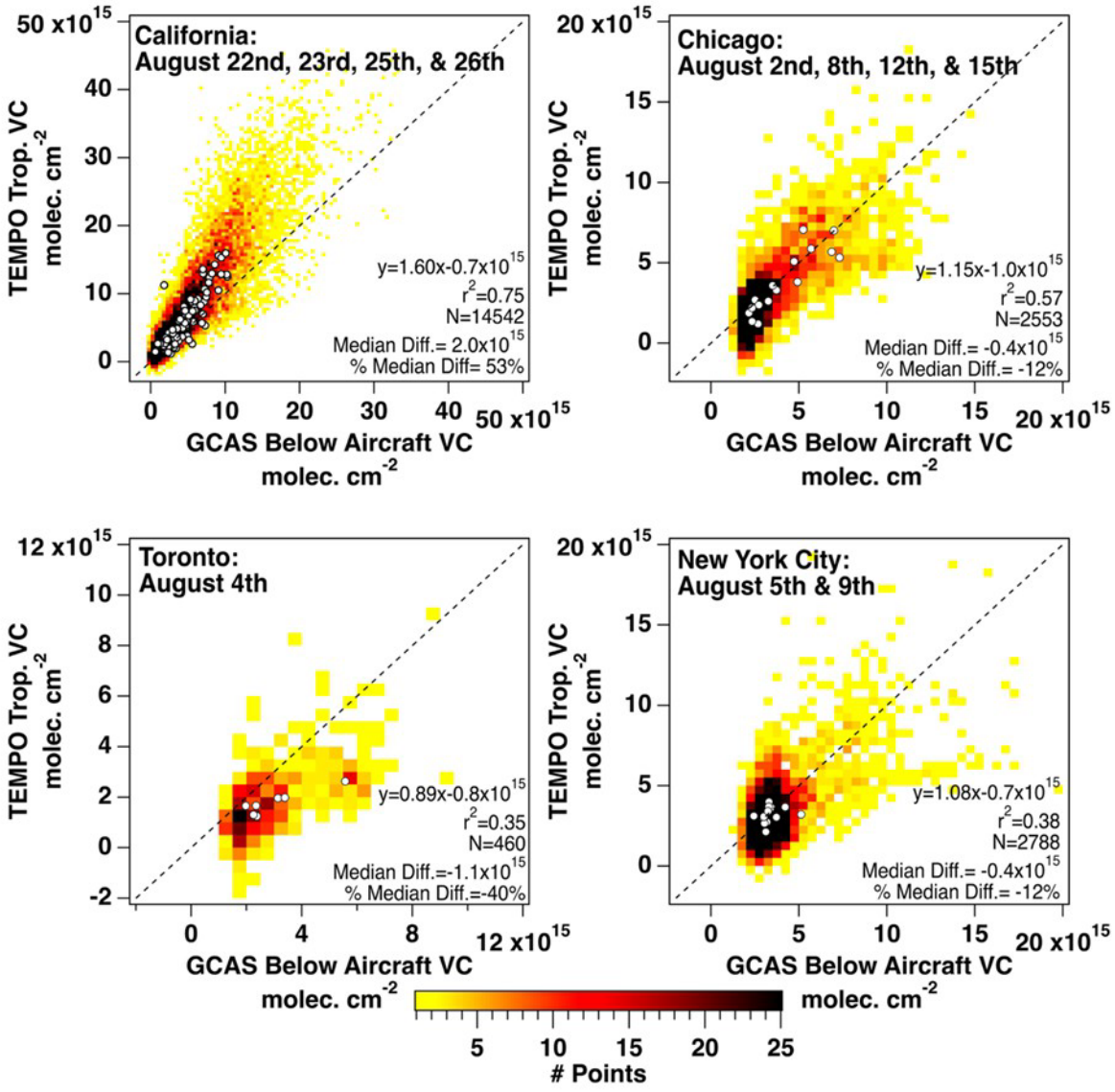


Figure 3-22: Density plots of GCAS Below Aircraft Vertical Column (VC) and the TEMPO Tropospheric Vertical Column for all cloud-free coincidences from STAQS GCAS flights separated by region flown. Noted are the dates flown in each region and statistics represent data within the density plot. Overlaid are the median values and interquartile ranges of each parameter over an entire TEMPO scan. Number of pixels range from 2 to 341 with an average of 210 and standard deviation of 109.

3.2.6 TropNO2 Inter-Comparison Between TEMPO and Airborne Multi-Axis DOAS measurements during CUPiDS

Authored by: Sunil Baidar (NOAA-CSL, CU Boulder), Rainer Volkamer, Christopher F. Lee, Rebecca Mesburis, and Nicole Silver (CU Boulder), Funding provided by NOAA AC4 award NA21OAR4310139

Dataset: NO_2 vertical column densities (VCDs) were measured by the University of Colorado Airborne Multi-Axis Differential Optical Absorption Spectroscopy (CU AMAX-DOAS) instrument aboard the NOAA Twin Otter aircraft during the Coastal Urban Plume Dynamics Study (CUPiDS) in the New York metro area

in July-August 2023. CU AMAX-DOAS is a UV-vis instrument, which is optimized to obtain tropospheric V CDs of trace gases as well as detailed profile information for trace gases and aerosol extinction (Baidar et al. 2013). The field of view of the instrument is 5.89° x 0.3° which translates to ~300 m x ~15 m footprint at the surface while flying at 3 km altitude. NO_2 differential slant column densities (dSCDS) are retrieved in the wavelength region of 425 to 490 nm using a zenith reference spectrum, obtained at altitude above 2 km AGL, outside the planetary boundary layer from the same flight. Nadir NO_2 dSCDs can be directly interpreted as tropospheric SCD (tropSCD) assuming the zenith reference only contains stratospheric NO_2 and that change in stratospheric NO_2 is negligible during the Twin Otter flights which usually lasted 3-4 hours during daytime (solar zenith angle < 65°). Changes in stratospheric NO_2 are tracked throughout the flight duration with frequent zenith measurements and corrected for higher solar zenith angle measurements using a polynomial fit. Small amounts of NO_2 exist between the flight altitude and the tropopause, but those are largely assumed to be static in flight. This is accounted for as described later in this section.

Nadir NO_2 dSCDs or tropSCDs are converted to tropospheric vertical column densities (tropVCDs) using geometric airmass factors (geoAMF),

$$tttttttVVVVVV = \frac{tttttttSSVVVV}{ggggttAAAAAA} = \frac{tttttttSSVVVV}{ssggcc\ ssggcc\ (SSSSAA) + 1}$$

GeoAMF assumes that all photons collected at the nadir geometry are reflected off the Earth's surface at the nadir point. Thus, the slant column is related to vertical column by secant of the solar zenith angle (SZA) plus the unity for light path from the nadir point to the instrument telescope. Note that aircraft height and hence light path enhancement is implicitly included by applying the trigonometric function of a right-angled triangle. The use of the geometric AMF makes the tropospheric NO_2 VCD independent of NO_2 profile assumptions, and other ancillary information needed to calculate AMFs using radiative transfer model simulations. The error due to the use of geometric AMF depends on the SZA and is expected to be < 6% for SZA < 20°, <10% for SZA < 50° and < 20% for SZA < 60° based on comparison with RTM simulations for predominant conditions encountered during the CalNex/CARES campaigns (Baidar et al., 2013). CU AMAX-DOAS measured NO2 tropVCDs have previously been used to evaluate NASA OMI tropospheric NO2 product (Oetjen et al., 2013).

During CUPiDS, AMAX-DOAS measurements included 24 coincident measurements (within ±5 km and ±15 minutes) with Pandora direct sun observations at 7 locations. This resulted in a $R^2 = 0.83$ and a slope of 0.94 and a small negative offset of $-7x10^{14}$ molecules cm^{-2}. The mean bias (Pandora – AMAX-DOAS) was $1.1x10^{15}$ molecules cm^{-2} and small part of it is likely due to the tropospheric NO_2 column above the aircraft present in the reference spectrum. For comparisons shown below, the NO_2 vertical column between the flight altitude and tropopause is calculated using TEMPO's *a priori* NO_2 profile (GEOS-CF) for each pixel and added to the corresponding AMAX-DOAS NO_2 data to represent a tropVCD.

Sampling strategy: AMAX-DOAS NO_2 VCD measurements collected above 2 km flight altitude and within ± 30 minutes of TEMPO measurements are compared with TEMPO tropospheric NO_2 VCDs. AMAX_DOAS data within each TEMPO pixel are averaged for the comparison. TEMPO tropospheric NO_2 VCD data is filtered for clouds (eff_cloud_fraction < 0.15) and has a main data quality flag equal to 0.

Metrics used: Orthogonal linear fits are applied to coincident data between the MAX-DOAS and TEMPO. A 3σ (ΔNO_2 VCD) filter is applied to the data before performing an orthogonal linear fit in order to reduce the impact of the outliers on the fit results. Correlation coefficient and mean bias (TEMPO – AMAX-DOAS) are also calculated. Statistics with and without the 3 σ filter are discussed below.

Results: This work helps in assessing PSPIs NO2-01and NO2-07.

Figure 3-23 illustrates how both TEMPO and AMAX-DOAS can distinguish NO_2 gradients over the New York metro area from polluted city centers to cleaner rural and coastal areas. The timeseries plot in the bottom panel shows how TEMPO captured the same NO_2 plumes along the AMAX-DOAS flight track over the 3 different hours during the duration of the flight. Although TEMPO showed larger pixel-to-pixel variability, it captured the same gradient and evolution of the NO_2 plumes over the region and hence satisfying **PSPI NO2-01**. AMAX-DOAS measurement peaks are also higher than TEMPO, but this is likely a reflection of the higher spatial resolution of the airborne measurement which would be averaged out over the TEMPO pixel area.

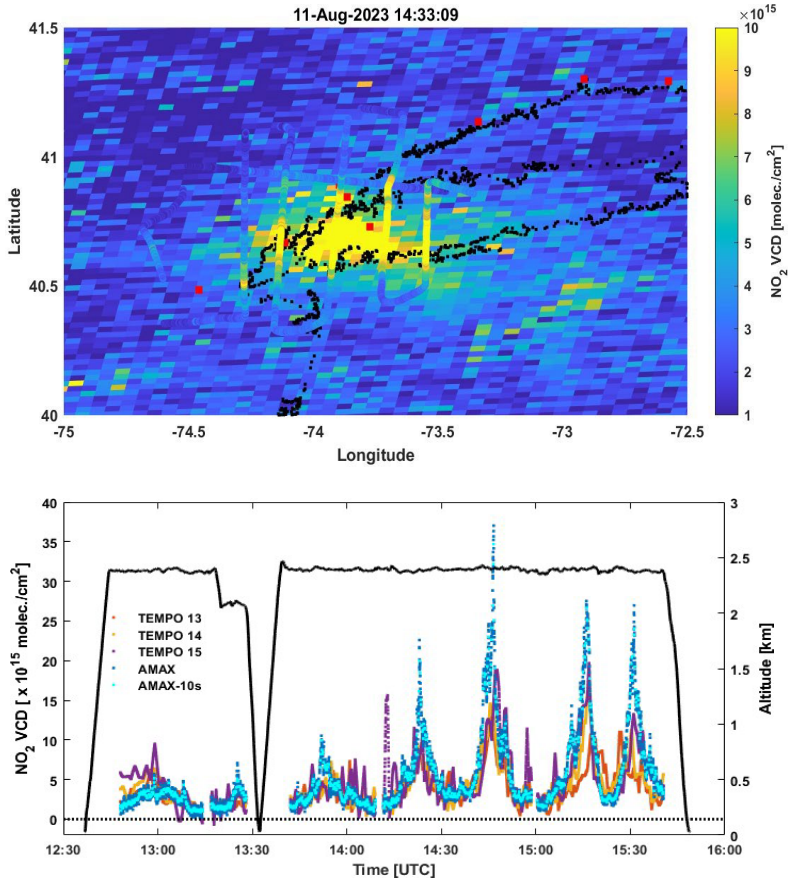


Figure 3-23: (top) Map of TEMPO NO2 VCDs measured on August 11, 2023, at 14:33 UTC. Overlaid on top is the NO2 VCD measured by AMAX-DOAS between 12:45-15:40 UTC. (bottom) Time-spatial series NO2 VCD measured by AMAX-DOAS and TEMPO at ~1330 UTC (orange), ~1430 UTC (yellow), and ~1530 UTC (purple).

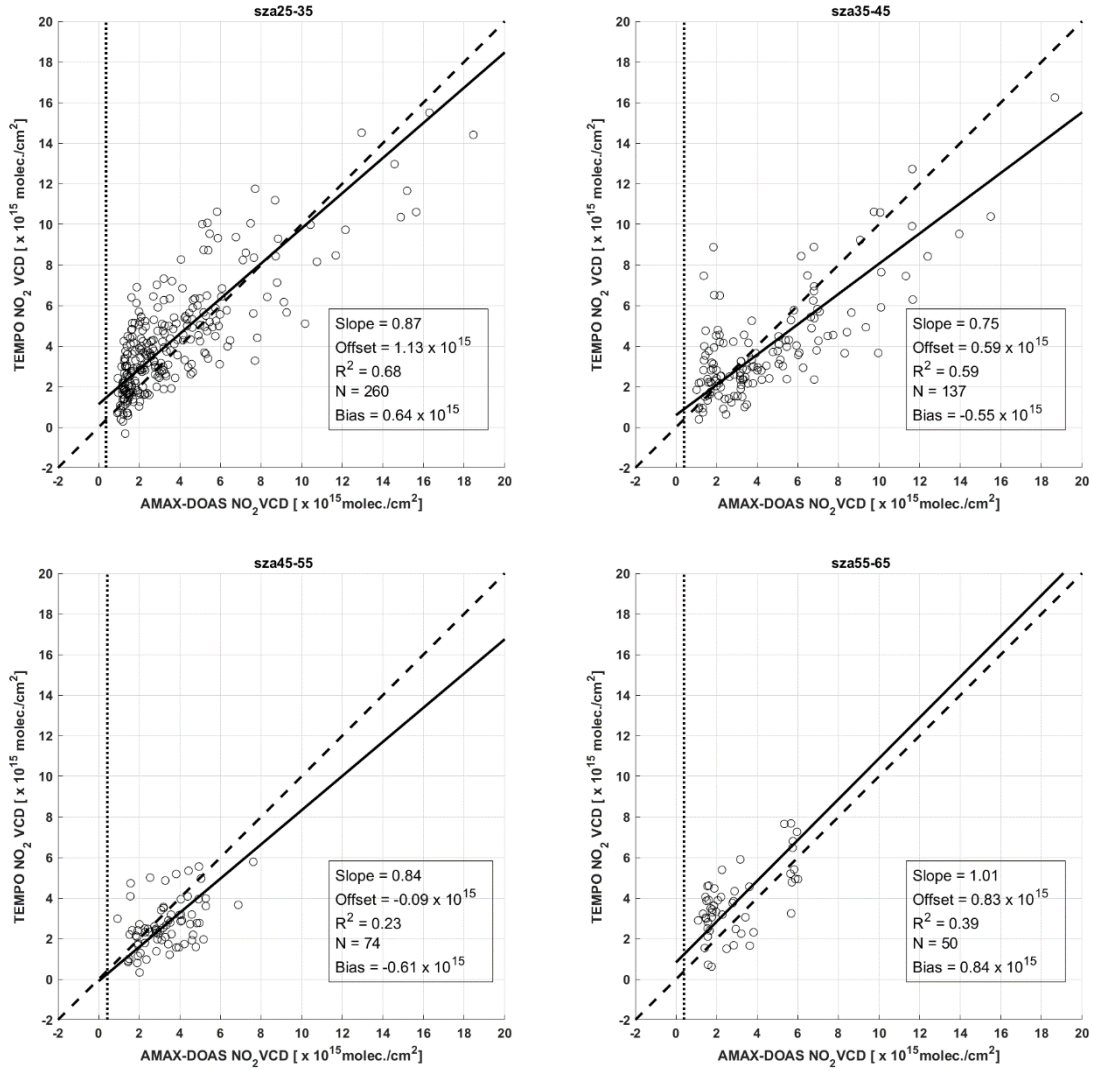


Figure 3-24: Scatterplot comparing TEMPO and AMAX-DOAS tropospheric NO2 VCD for different solar zenith angles bands: (a) 25-35 degrees, (b) 35-45 degrees, (c) 45-55 degrees, and (d) 55-65 degrees. The dotted vertical line represents the average NO2 VCD, between the aircraft flight altitude and tropopause calculated from TEMPO a priori NO2 profiles, which is added to the AMAX-DOAS measurements. The solid black line is the orthogonal fit to the data. The dashed black line represents the 1:1 line.

Figure 3-24 shows scatterplots comparing TEMPO and AMAX-DOAS NO2 tropospheric VCD for different solar zenith angle (sza) ranges and quantitative results supporting **PSPI NO2-07**. Each AMAX data point represents a mean NO_2 VCD measured within a TEMPO pixel and number of data points ranged from 1-108 with a median of 34 measurements. The dotted vertical line represents the average NO_2 VCD between the aircraft flight altitude and tropopause that was added to AMAX-DOAS below aircraft measurements (~0.2 x 10^{15} molecules cm^{-2}). Most data points span the 1:1 line (dashed). Although the R^2 values exhibit lower correlations at higher sza bands due to limited dynamic range of the data, the slopes (0.75-1.01) show good agreements between the two data sets There does not appear to be any major systematic differences in TEMPO data quality with solar zenith angle, although limited number of data points and

dynamic range of the data especially at higher solar zenith angles makes it difficult to draw robust conclusions. Combined analysis of all 4 solar zenith angle bands shows that there is very good agreement (r^2= 0.59, slope of 0.83) between TEMPO and AMAX-DOAS datasets collected during CUPiDS (see Figure 3-25). There is a small positive bias of 0.18 x 10^{15} molecules cm^{-2}, which is much smaller than the bias observed between AMAX-DOAS and Pandora (1.1x10^{15} molecules cm^{-2}). Not applying a 3σ ΔNO2 VCD filter to the complete dataset results in no change in r^2, but an increase in bias to 0.36x10^{15} molecules cm^{-2}, slope change to 0.65 and offset of 1.71x10^{15} molecules cm^{-2}.

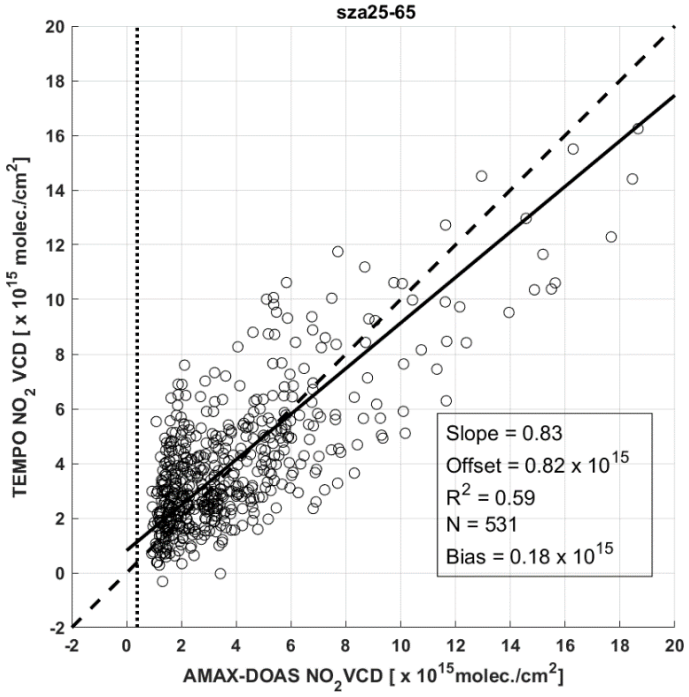


Figure 3-25: Same as Figure B with all 4 sza bands combined together.

3.2.7 TropNO2 Evaluation of TEMPO evaluation using WRF-Chem

Authored by: R. Bradley Pierce, Jerrold Acdan, Maggie Bruckner, (UW-SSEC); Laura Judd, Scott Janz, (NASA); Mike Newchurch, (UAH-MSFC); Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO-TEMPO)

Chemical transport models (CTMs) can play an important role in TEMPO validation activities by providing a means of including non-coincident measurements in the validation process. This is referred to as "indirect validation" (Zhu et al, 2020) and is accomplished by:

1) Comparison between the CTM and non-coincident validation measurements to determine CTM biases and RMSE
2) Comparison between the CTM and satellite retrievals to determine the biases between the CTM and retrievals.
3) Comparison of these two sets of CTM biases and RMSE can be used to indirectly assess the biases and RMSE between satellite and the non-coincident validation measurements.

We use STAQS GCAS remote sensing NO2 measurements to evaluate high resolution WRF-Chem simulations during August 2023 over Chicago (4 days), New York (3 days), and Toronto (1 day). We then compare the WRF-Chem simulations to TEMPO V3 NO2 retrievals for the same regions:

- Chicago (11 days, 86 Granules, 12Z-23Z) G04, G05, G06 [41N<lat<43N, 88.5W<lon<87W]
- NYC (11 days, 89 Granules, 12Z-23Z) G02, G03, G04 [40N<lat<42N, 74.5W<lon<72W]
- TOR (11 days, 79 Granules, 12Z-23Z) G02, G03, G04 [43N<lat<44N, 80W<lon<79W]

This indirect validation approach allows us to evaluate the TEMPO NO2 column retrieval under a wider range of conditions (11 days) compared to 1-4 days using GCAS direct validation. Indirect validation also allows us to evaluate diurnal and urban/suburban variations in the TEMPO NO2 column.

We use the following filtering criteria for the TEMPO data: QC=0, cloud fraction less than or equal to 0.15, and Column NO2 greater than 0.0. We also apply box-car averaging (width=2, Equation 3.2.7.1) to remove pixel-to-pixel variability in both the along-track (xtrack) and cross-track (mirror_step) directions for each TEMPO granule.

$$RR_{ii=\{\frac{1}{ww}\sum_{jj=0}^{ww-1} AA_{iiii^{ii+jj-\frac{ww}{2}}}} \quad \frac{ww-1}{2} \le ii \le NN - \frac{ww+1}{2} \qquad (EEEE.\ 3.2.7.1)$$

Table 3-3 describes the WRF-Chem modeling framework.

Domains	4-km centered on Chicago and New York City
Vertical Levels	40 levels with high resolution below 2km
Meteorology	HRRR 3-km
Chemical Boundary and Initial Conditions	WACCM with Canadian wildfires scaled by physiographic region, QFED fire detections with FINNv1.5 emission ratios applied
Anthropogenic Emissions	NEMO 1-km based on NEI17, non-trended, upscaled to 4-km
Biogenic Emissions	MEGAN v2
Surface Physics	Noah-MP LSM
Chemistry Mechanism	T1_MOZCART (updated MOZART chemistry, GOCART aerosols)

Table 3-3 WRF-Chem Retrospective continuous modeling framework. Acronyms are HRRR: High-Resolution Rapid Refresh, QFED: Quick Fire Emissions Dataset, FINN: Fire INventory from NCAR, NEMO Neighborhood Emission Mapping Operation

Figure 3-26 shows scatter plots of the WRF-Chem predictions and the STAQS GCAS below aircraft NO2 column for each of the three urban areas (Chicago, Toronto, and New York City) sampled during the Eastern Phase of the NASA STAQS campaign (July 26 through August 15, 2023). The WRF-Chem NO2 columns are well correlated with GCAS within each urban area (r=0.77 to 0.79) with biases ranging from.

molecules cm^{-2}. These comparisons establish the overall biases and RMSE between the validation data (remotely sensed NO2 from GCAS) and the WRF-Chem CTM.

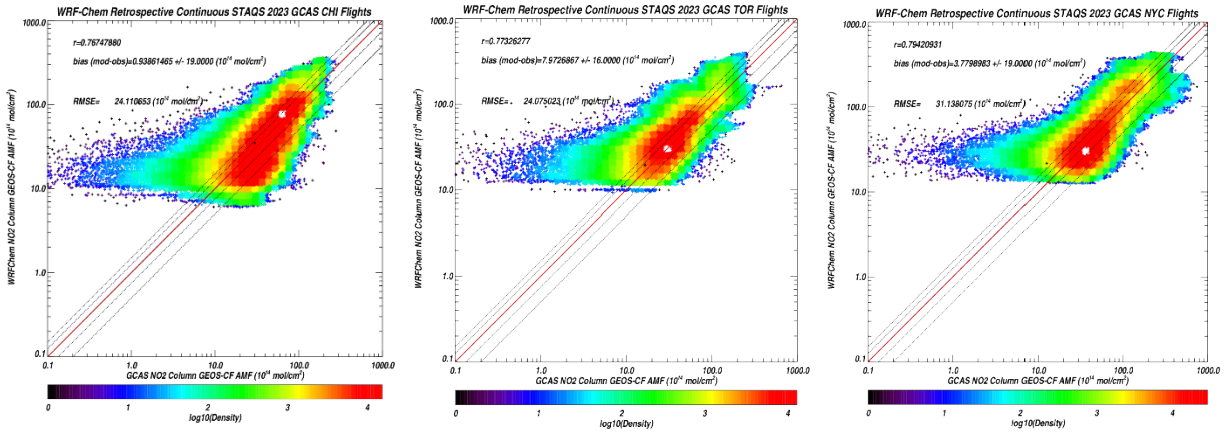


Figure 3-26 Figure 3.2.7.1: Scatter plots of WRF-Chem verses GCAS column NO2 for Chicago (CHI, left), Toronto (TOR, middle) and New York City (NYC, right). GCAS column NO2 was provided by L. Judd. Solid red line in the scatter plot is the 1-to-1 line, dashed lines show 25% and 50% errors.

Figure 3-27 shows scatter plots of the WRF-Chem predictions and the TEMPO V3 box-car averaged column NO2 within the regional boxes for all TEMPO observations over each of these cities. The WRF-Chem NO2 columns are reasonably well correlated with TEMPO within each urban area (r=0.63 to 0.73) with biases ranging from $9.8x10^{14}$ +/- $8x10^{14}$ to $14.7x10^{14}$ +/- $17x10^{14}$ molecules cm^{-2} and RMSE ranging from $22x10^{14}$ to 27.5 $x10^{14}$ molecules cm^{-2}.

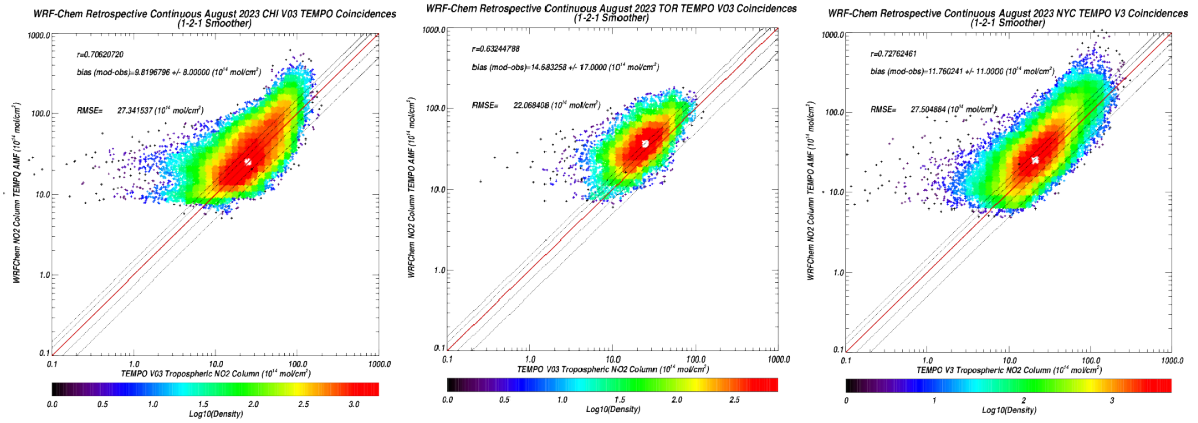


Figure 3-27: Scatter plots of WRF-Chem verses TEMPO V3 column NO2 for Chicago (CHI, left), Toronto (TOR, middle) and New York City (NYC, right). Solid red line in the scatter plot is the 1-to-1 line, dashed lines show 25% and 50% errors.

By taking the difference between the WRF-Chem - GCAS (Figure 3-26) and WRF-Chem - TEMPO (Figure 3-27) biases we can infer the TEMPO biases relative to GCAS; . This results in inferred TEMPO-GCAS biases of -$8.9x10^{14}$ +/- $21.x10^{14}$, -$6.7x10^{14}$ +/- $23.x10^{14}$, and -8.0 $x10^{14}$ +/- $22.x10^{14}$ molecules cm^{-2} for Chicago, Toronto, and New York City, respectively. The RMSE between WRF-Chem and TEMPO are very consistent with the RMSE between WRF-Chem and GCAS and are 2-3 times higher than the TEMPO NO2 column precision requirement of $10x10^{14}$ molecules cm^{-2}. Some of this RMSE comes from uncertainties in the WRF-Chem simulations and so this is likely an overestimate of the true TEMPO RMSE.

To evaluate urban/suburban and diurnal variations in the TEMPO NO2 column retrievals we sub-set the WRF-Chem simulations and TEMPO V3 NO2 retrievals into “urban” and ‘suburban” samples for the same regions used in the previous analysis. The sub-regions are defined as:

- Chicago (11 days, 86 Granules, 12Z-23Z) G04, G05, G06
  - Chicago [41.65N<lat<41.9N, 87.95W<lon<87.7W]
  - Northwestern Suburbs [42.2N<lat<42.5, 88.45W<lon<88.2W]
- NYC (11 days, 89 Granules, 12Z-23Z) G02, G03, G04
  - NYC [40.6N<lat<40.8N, 74.25W<lon<73.75W]
  - CT [41.4N<lat<41.6N, 72.75W<lon<72.25W]

We then bin the coincident TEMPO and WRF-Chem data into hourly means during the Eastern Phase of the NASA STAQS campaign. We also include the TEMPO GEOS-CF *a priori* in these evaluations. Figure 3-28 shows the results of this evaluation. The mean TEMPO NO2 column does not show a strong diurnal variation for any of the regions considered. The GEOS-CF *a priori* shows a stronger diurnal variation, particularly in urban New York City where the morning (12-14Z) is higher than the rest of the day. The WRF-Chem simulated NO2 column (with the TEMPO airmass factor and scattering weights applied) is generally slightly higher than TEMPO throughout the day but in better agreement with the TEMPO NO2 column than the GEOS-CF *a priori* is. The TEMPO NO2 column retrieval is able to distinguish between urban and neighboring suburban regions in both Chicago, where the TEMPO NO2 column is 50% lower in the Northwestern Suburbs (2.42 x10^{14} molecules cm^{-2}) compared to urban Chicago (4.86 x10^{14} molecules cm^{-2}), and New York City, where the TEMPO NO2 column is 70% lower in Connecticut (1.85e x10^{14} molecules cm^{-2}) compared to urban New York City (6.18 x10^{14} molecules cm^{-2}). Both suburban regions also show less variability compared to their urban counterparts.

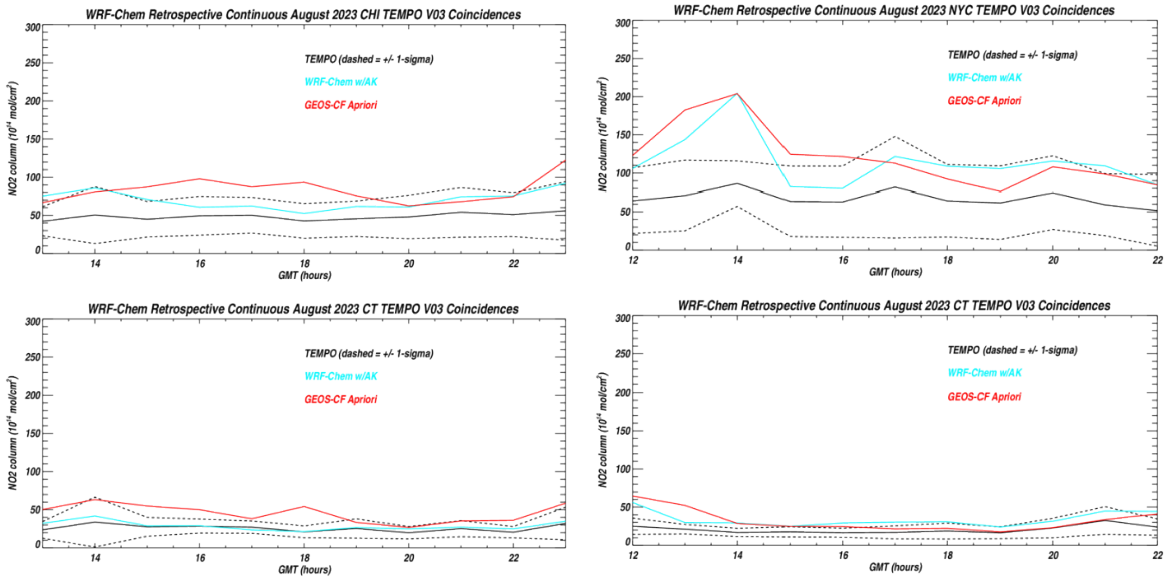


Figure 3-28: Mean diurnal variations in TEMPO (black), WRF-Chem (blue), and the GEOS-CF a priori (red) NO2 columns during August 2023 for Chicago urban (CHI, upper left), New York City urban (NYC, upper right), Chicago Western Suburbs (Western Suburbs, lower left), and Connecticut (CT, lower right). The dashed black lines show +/- one standard deviation (1-sigma) for the TEMPO NO2.

3.2.8 TEMPO GEOS-CF a priori evaluation

Authored by: Viral Shah (NASA GSFC), Sungyeon Choi (NASA GSFC), Lok N. Lamsal (NASA GSFC, now at Bureau of Ocean Energy Management; BOEM), Matthew S. Johnson (NASA Ames), Claudia Bernier (NASA Ames)

The NASA Goddard Earth Observing System Composition Forecast (GEOS-CF v1.0) modeling system provides global high-resolution (approx. 25×25 km^2) near real time estimates and forecasts of atmospheric composition (Keller et al., 2021; Knowland et al., 2022). It uses the GEOS-Chem atmospheric chemistry model embedded within the NASA Global Modeling and Assimilation Office's (GMAO's) GEOS Earth System Model (ESM) and builds on scientific advances in atmospheric chemistry provided by the GEOS-Chem community as well as on the state-of-the-art modeling capabilities at the GMAO. GEOS-CF provides the a priori vertical profiles and other model-derived information (e.g., temperature profiles and tropopause pressure) for the TEMPO NO_2 and HCHO column retrievals. Here we check the consistency between the TEMPO NO_2 retrievals (V03) and the GEOS-CF NO_2 columns and evaluate the GEOS-CF NO_2 vertical profiles using aircraft observations from the Atmospheric Emissions and Reactions Observed from Megacities to Marine Areas (AEROMMA) campaign.

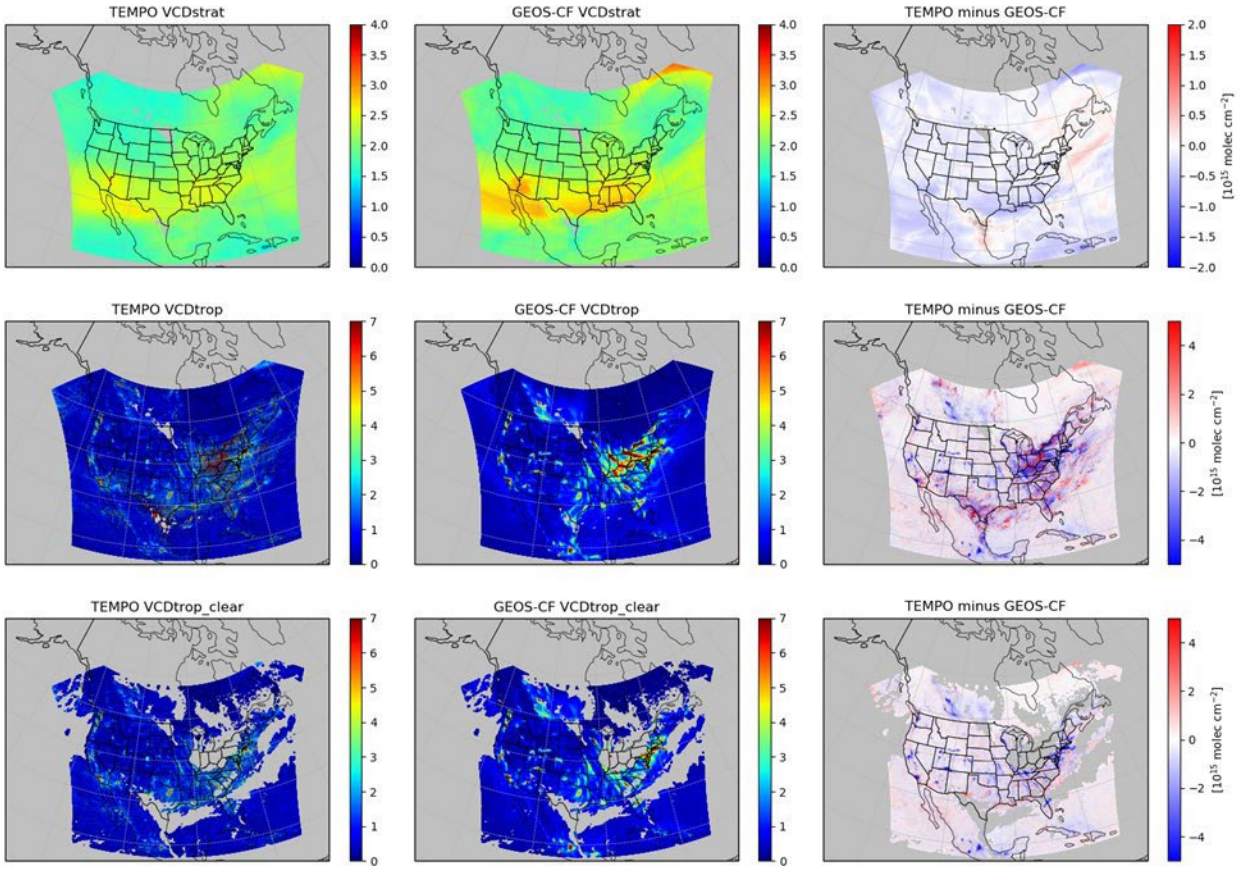


Figure 3-29: Comparison of TEMPO-retrieved and GEOS-CF stratospheric (top) and tropospheric (middle and bottom) NO2 columns during 11th scan in early afternoon (18 UTC) on March 18, 2024. The middle panels show all data from the scan, while the bottom panels retain only clear-sky and partially cloudy conditions based on the TEMPO cloud retrieval (eff_cloud_fraction<0.3).

Figure 3-29 shows TEMPO Level 2 stratospheric and tropospheric NO_2 column retrievals compared with the coincidently sampled GEOS-CF simulated NO2 columns for one early-afternoon scan on March 18, 2024. In general, the distribution of NO_2 is similar between TEMPO and GEOS-CF with higher values in the stratosphere and lower values in troposphere except over highly polluted areas. Stratospheric NO_2 columns are highly correlated, and the enhanced stratospheric features are observed in both. However, TEMPO stratospheric NO_2 columns are generally higher than GEOS-CF values for the scan shown in Figure 3-29. In the troposphere, TEMPO and GEOS-CF are fairly well correlated and the GEOS-CF simulation tends to be higher than TEMPO observations. Differences are larger and retrievals are noisier over cloudy scenes, where retrieval uncertainty is expected to be higher.

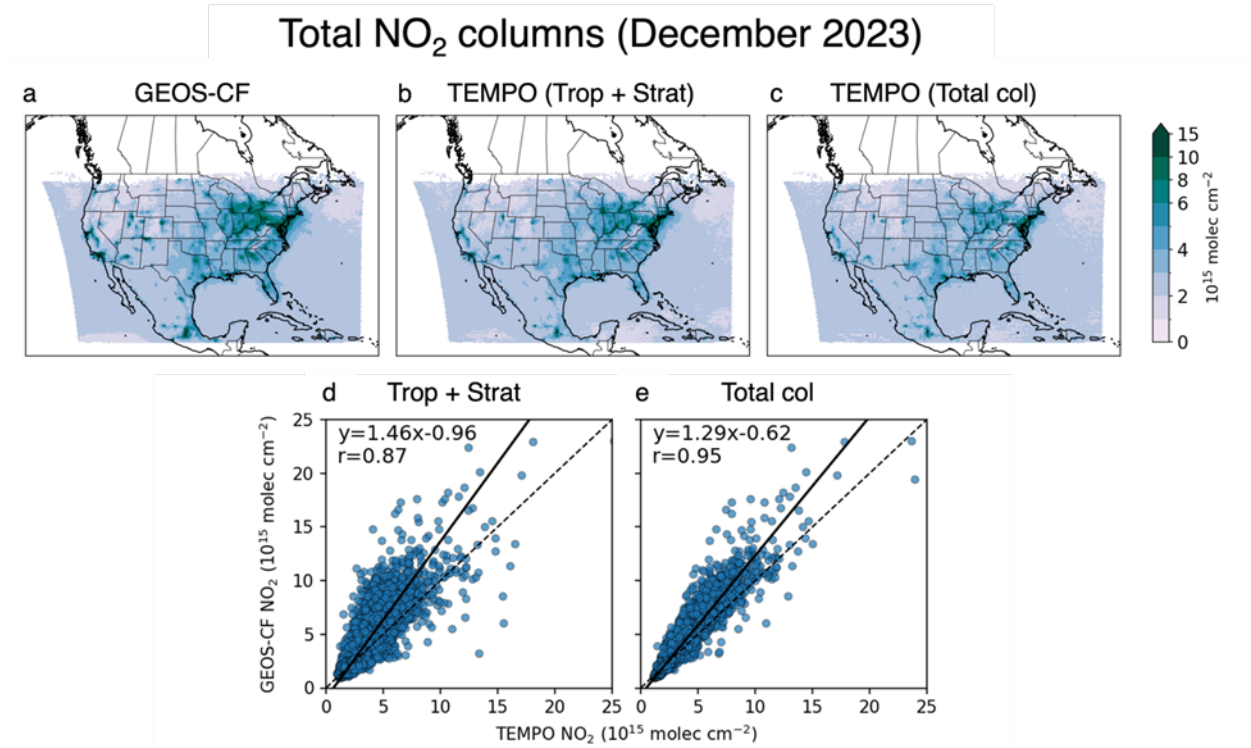


Figure 3-30: Comparison of GEOS-CF and TEMPO NO2 total vertical columns for December 2023. The top panels show the monthly mean (a) GEOS-CF total columns, (b) sum of tropospheric and stratospheric columns retrieved by TEMPO (vertical_column_troposphere + vertical_column_stratoshere), and (c) total columns retrieved by TEMPO (vertical_column_total). The TEMPO Level 2 data are gridded to the GEOS-CF grid (0.25° lat × 0.25° lon). Only TEMPO pixels with good quality retrievals (main_data_quality_flag = 0), clear-sky and partially cloudy conditions (eff_cloud_fraction<0.3), and solar zenith angles less than 70° are used. GEOS-CF is sampled at the same time as the TEMPO observations. The bottom panels show scatter plots of the GEOS-CF columns versus (a) sum of TEMPO tropospheric and stratospheric columns and (b) TEMPO total columns. The text inset shows the linear regression slopes and intercept and the correlation coefficient.

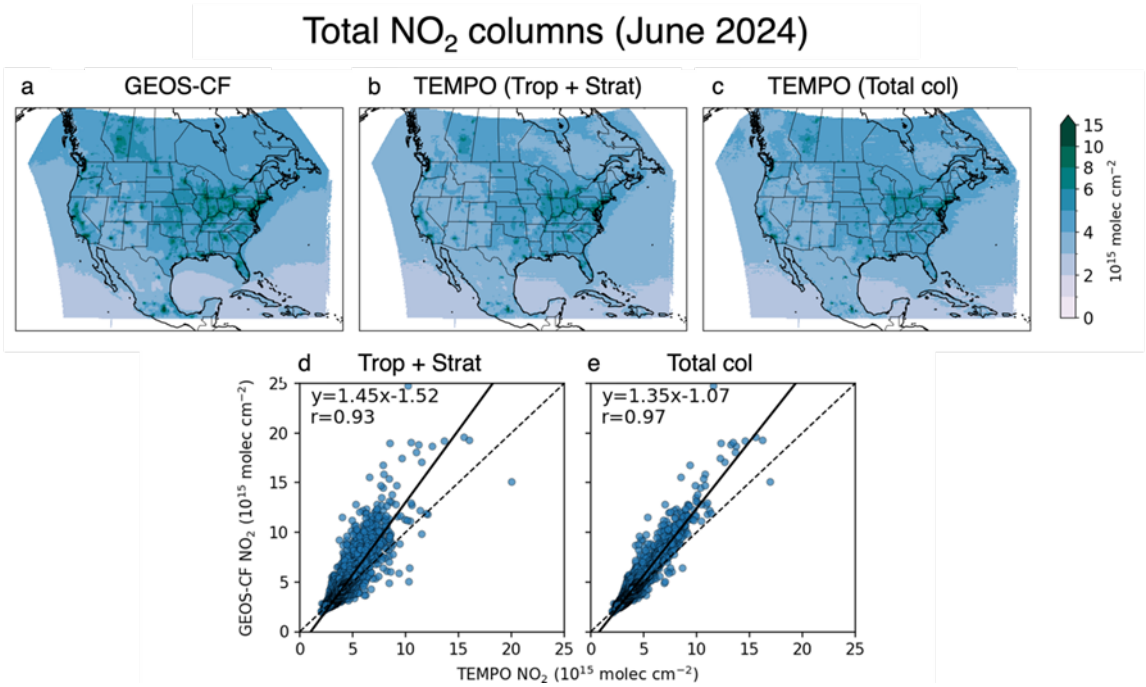


Figure 3-31: Same as Fig. 3-30 Fig. 3.2.8.2 but for June 2024.

Figures 3-30 and 3-31 compare daytime average of total NO_2 vertical columns from TEMPO and GEOS-CF for December 2023 and June 2024, respectively. We compare the GEOS-CF columns to both the sum of the stratospheric and tropospheric (trop + strat) TEMPO columns calculated using the stratosphere-troposphere separation method and the TEMPO total NO_2 columns, which are calculated by dividing the total slant columns with the air mass factor for the entire atmosphere. While both provide estimates of the total NO_2 vertical columns, the trop + strat NO_2 columns are recommended for total column comparisons (TEMPO User Guide). Both the trop + strat and the total TEMPO NO_2 columns replicate the spatial patterns seen in the GEOS-CF columns with a high correlation, partly reflecting the influence of the GEOS-CF a priori profiles on the TEMPO NO2 retrievals. There is more scatter in the TEMPO trop + strat columns and GEOS-CF comparison, likely because the total column retrievals use information from the GEOS-CF a priori profiles in the stratosphere as well as the troposphere. The GEOS-CF NO_2 columns are on average 29–46% higher than the TEMPO columns, which may be due to an overestimate in NOx emissions in the model.

Figure 3-32 compares the tropospheric NO_2 vertical columns from GEOS-CF and TEMPO. Similar to Figs. 3-29 – 3-31, we see high spatial correlation between the GEOS-CF and TEMPO fields, with GEOS-CF fields showing a high bias (~50%), again likely reflecting an overestimate in NOx emissions in the model. However, over western Texas, TEMPO shows higher NO_2 columns than GEOS-CF, particularly in December, suggesting missing oil and gas emissions in the emissions inventory used in GEOS-CF.

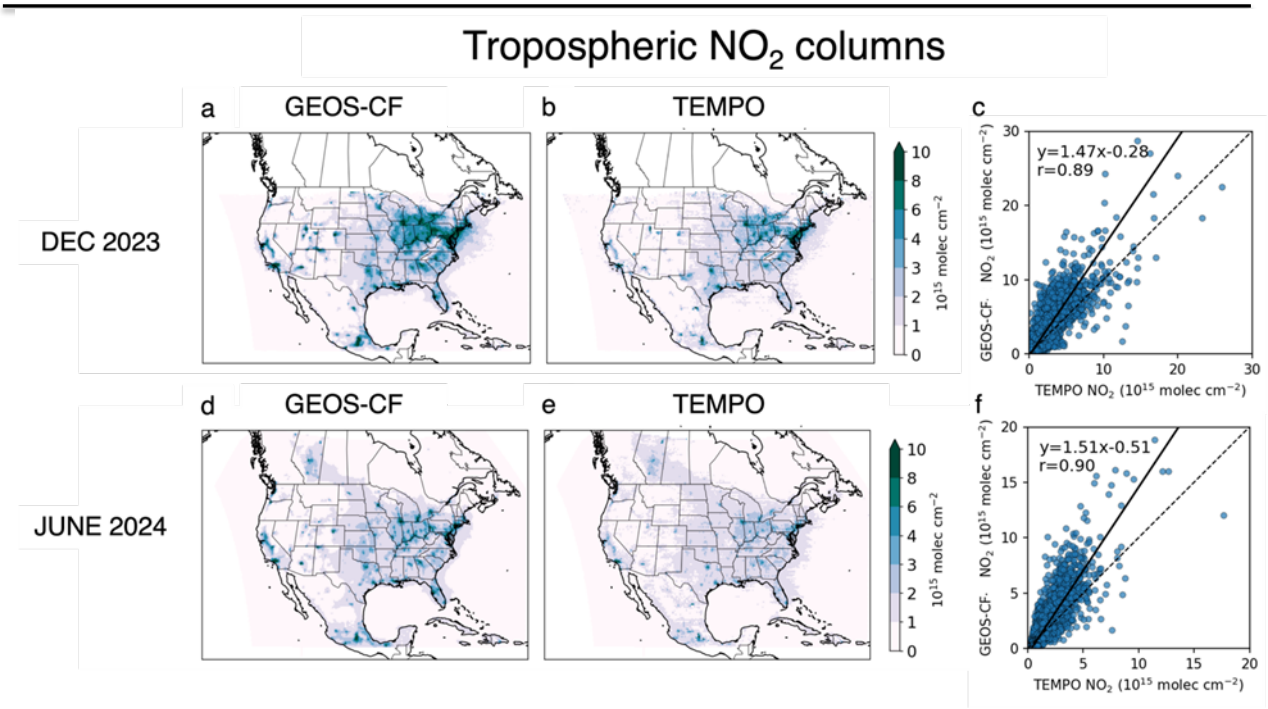


Figure 3-32: Comparison of GEOS-CF and TEMPO NO2 tropospheric vertical columns for December 2023 and June 2024. Figures show the daytime monthly mean maps of GEOS-CF (a,d) and TEMPO (b,e) tropospheric columns (vertical_column_troposphere), and the scatter plots of the GEOS-CF versus the TEMPO tropospheric columns (c,f). The TEMPO data selection and processing are described in caption of Fig. 3-30. The text inset in the scatter plots shows the linear regression slopes and intercept and the correlation coefficient.

Figure 3-33 compares the GEOS-CF and TEMPO stratospheric NO_2 columns. Both the GEOS-CF and TEMPO stratospheric NO_2 columns are higher in June than in December because of faster photolysis of NOx reservoir species (HNO_3 and N_2O_5) and faster oxidation of N_2O. We find good spatial correlation between the GEOS-CF and TEMPO columns (r = 0.95), with the GEOS-CF stratospheric columns slightly higher than the TEMPO retrievals in June, consistent with Fig. 3-29. The bottom panels of Fig. 3-33 show the hourly variation in stratospheric NO_2 columns in six longitude bands. There is an increase in NO_2 columns in GEOS-CF of 0.5–1×10^{15} molecules cm^{-2} during the day because of photolysis of HNO_3 and N_2O_5, which is also observed in the TEMPO data.

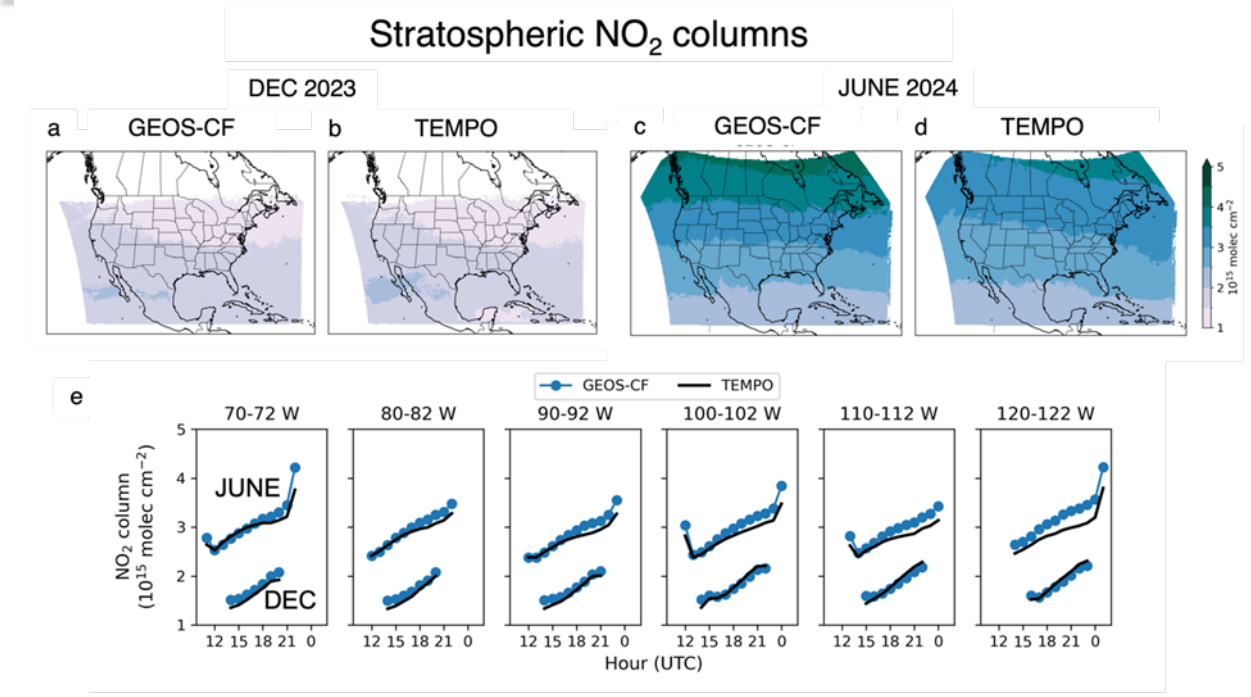


Figure 3-33: Comparison of GEOS-CF and TEMPO NO2 stratospheric vertical columns for December 2023 and June 2024. The top panels show the monthly mean GEOS-CF (a,c) and TEMPO (b,d) stratospheric columns (vertical_column_stratosphere). The bottom panels show the hourly variation of the GEOS-CF and TEMPO stratospheric columns in 6 longitude bands, from east to west.

In order to understand the potential biases in TEMPO NO_2 vertical column retrievals from the a priori shape provided by GEOS-CF model simulations, we compared simulated values to Airborne Cavity Enhanced Spectrometer (ACES) NO_2 observations from DC-8 flights during the AEROMMA summer 2023 campaign. To avoid representation errors, all ACES observations on the DC-8 flights were averaged to match the spatial (0.25° × 0.25°) and temporal (hourly) resolution of the GEOS-CF model. A preliminary focus was on the capability of GEOS-CF to reproduce the differences between observed NO2 vertical profiles in urban versus rural regions. Figure 3-34 shows the averaged comparison of GEOS-CF NO2 simulated concentrations compared to airborne observations during the AEROMMA campaign. This figure shows the model has moderate correlation with observations in both urban and rural areas; however, GEOS-CF displays larger biases and RMSE values for urban regions compared to rural areas. The model-data comparison suggests a large positive bias in GEOS-CF for the lowermost portion of the troposphere (0-2 km agl) in urban regions.

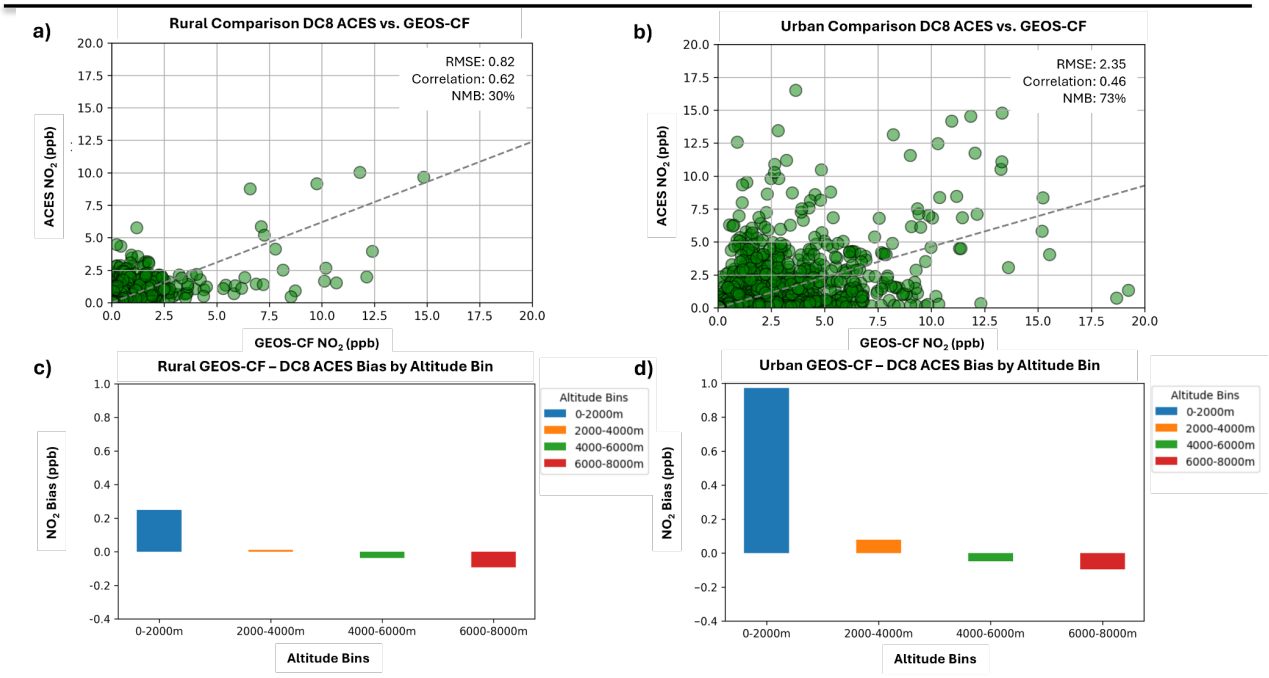


Figure 3-34: GEOS-CF NO2 concentrations (ppb) compared to co-located ACES DC-8 flight data (ppb) during the summer 2023 AEROMMA campaign. Scatter plots of the comparison are shown for rural (a) and urban (b) regions of the US. GEOS-CF NO2 biases for 2 km vertical bins for rural (c) and urban (d) regions are also displayed. Statistics for NMB (%), root mean squared error (RMSE, ppb), and correlation are shown in the figure inset. Urban and rural regions are separated using GEOS-CF NO emissions fields using a threshold of 2 x 10^{-10} kg m^{-2} s^{-1}.

Campaign-averaged vertical profiles from GEOS-CF and DC-8 ACES NO2 observations are displayed for urban and rural areas in Fig. 3-34. This figure further emphasizes the high bias in GEOS-CF NO2 predictions in the planetary boundary layer (PBL) region of the troposphere. This high bias in the model is not as pronounced for rural regions. Above the PBL region, GEOS-CF tends to compare well to airborne observations displaying a small low bias in the free- to upper-tropospheric in both urban and rural areas. The differences in the ability of GEOS-CF to reproduce observed vertical NO2 profiles in urban/rural regions could lead to differences in TEMPO NO2 vertical column biases caused by these shape factor errors resulting in region-dependent AMF errors.

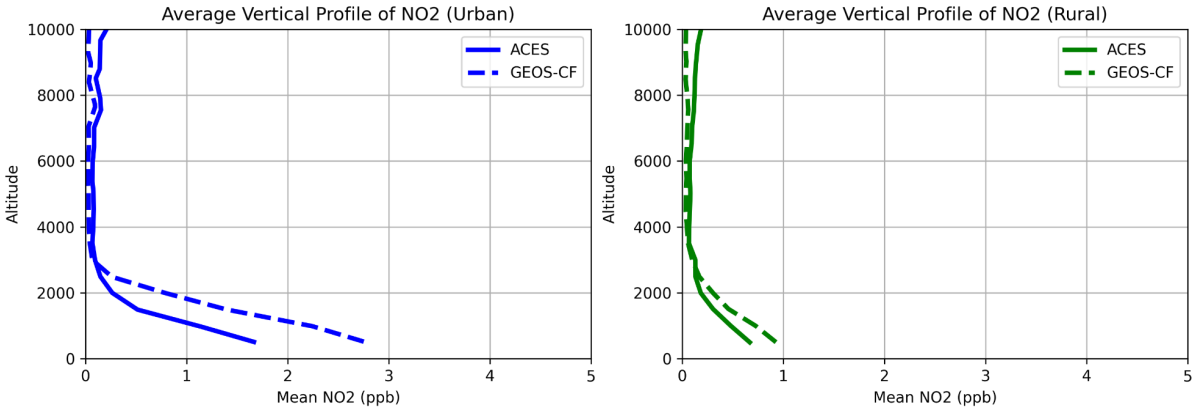


Figure 3-35: Mean GEOS-CF NO2 concentration vertical profiles (ppb, dashed lines) compared to co-located ACES DC-8 flight data (ppb, solid lines) during the summer 2023 AEROMMA campaign. The vertical profiles from the model and observations are shown for urban (left) and rural (right) regions. Urban and rural regions are separated using GEOS-CF NO emissions fields using a threshold of 2 x 10^{-10} kg m^{-2} s^{-1}.

3.2.9 TEMPO vs. TROPOMI Total NO_2 VCD comparison

Authored by: Ruijun Dang and Daniel Jacob (Harvard)

We compare TROPOMI NO2 with collocated TEMPO NO2 observations. We focus on NO2 total vertical column densities (VCDs).

Data descriptions

For TROPOMI NO2, we use the offline (OFFL) L2 product produced by the KNMI NO2 processor version 2.4.0 (https://sentinel.esa.int/web/sentinel/data-products/-/asset_publisher/fp37fc19FN8F/content/sentinel-5-precursor-level-2-nitrogen-dioxide). Only data with qa_value>0.75, cloud fraction<0.3, SZA<70°, surface albedo<0.3, and without snow/ice are used.

We conduct validations for approximately 1-month (20240513-20240616) of version 3 (V3) retrievals. Only data with main_data_quality_flag=0, cloud fraction<0.2, SZA<70°, surface albedo<0.3, and without snow/ice are used.

For the comparison, we select collocated TEMPO observation within a 10 km radius of the center of the TROPOMI pixel and within ± 30 min of the TROPOMI overpass time. TROPOMI retrievals are reprocessed using the same GEOS-CF NO2 a profile as in TEMPO. Pixel observations are gridded to a resolution of 0.25° × 0.3125° (~25 × 25 km^2). Only grid cells with more than 30 pairs of successful retrievals are included in the comparison.

Results

Figure 3-36 compares the spatial distribution of NO2 total VCDs from TROPOMI and TEMPO V3. Figure 3-37 shows the scatterplot of individual grid cells in Figure 1.

Overall, TEMPO TotNO2 is ~10% lower than TROPOMI. This low bias is mostly associated with background TotNO2 columns but is also evident over the most polluted urban areas, such as Los Angeles and Chicago. TEMPO shows a high bias relative to TROPOMI over other major urban areas, such

as Minneapolis and New York City. When these highly polluted grid cells (TROPOMI NO2 total columns>1x10^{16} molecules cm^{-2}) are excluded, R^2=0.96 and NMB=-11% (Figure 3-37). Note that this low bias relative to TROPOMI over highly polluted urban areas such as Los Angeles and New York City is consistent with previous TEMPO/TROPOMI comparisons (e.g. section 3.2.2) which found that TEMPO was lower than TROPOMI in Los Angeles and very close to TROPOMI in New York City.

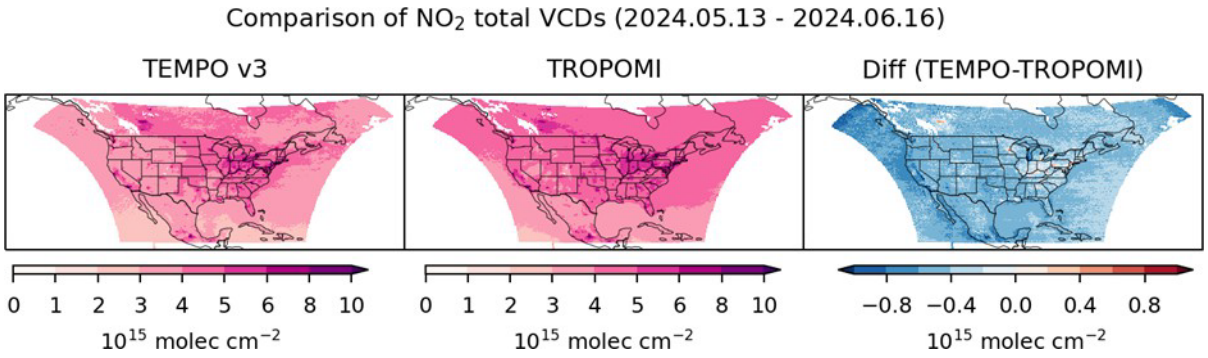


Figure 3-36: Spatial distribution of NO2 total columns from TEMPO version 3, TROPOMI, and the difference between the 2 products.

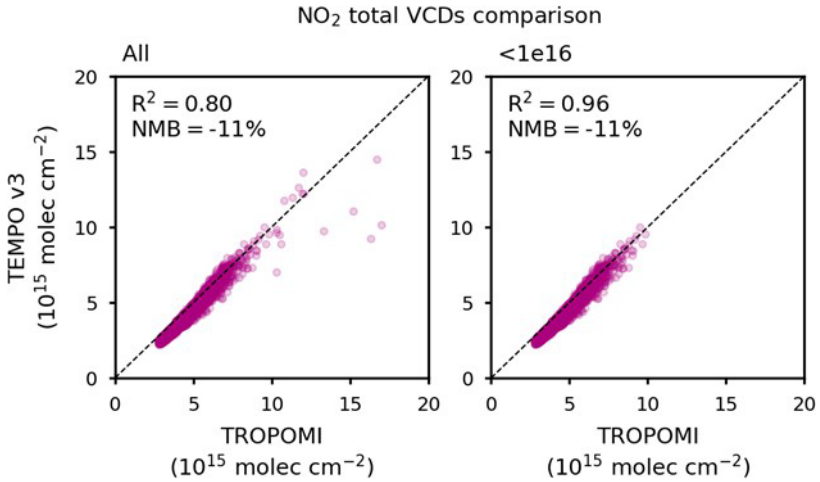


Figure 3-37: TEMPO v3 versus TROPOMI NO2 total columns at individual grid cells in Fig.3.2.9.1. The dashed line indicates the 1:1 line. R2 and normalized mean bias (NMB) are given inset. Left panel includes all grid cells, while right panel excludes grid cells with TROPOMI NO2 total columns greater than 1x10^{16} molecules cm^{-2}.

3.3 Level 2 Science Product Validation and Evaluation TEMPO_HCHO L2_V03 - TEMPO HCHO total column V03

3.3.1 TotHCHO Inter-Comparison Between TEMPO and TROPOMI

Authored by: : Barron Henderson, Lukas Valin, and Todd Plessel (EPA-ORD); Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO-TEMPO)

Introduction

This section compares TEMPO to the preexisting TROPOMI HCHO tropospheric product to address validation PSPIs HCHO-01, HCHO-02, HCHO-04, and HCHO-06. The results here have also been used to guide analyses for PSPI HCHO-05.

Methods

Datasets: TEMPO and TROPOMI total vertical column density (VCD) formaldehyde (HCHO) products are coincidently sampled within specific analysis regions. The TEMPO HCHO product that is the focus here is referred to as "version 3" that has data been reprocessed to cover the entire period of record (August 2023 to July 2025). TEMPO pixels are only used where the effective cloud fraction is less than 0.15 and QA flags are normal. The TROPOMI product discussed here is the "offline" version that is processed approximately 2-weeks after the measurements are made (S5P_L2__HCHO___HiR; C1918210023-GES_DISC). The TROPOMI HCHO product version varies slightly over the period of record. The version started at 2.5.0 in August 2023 and ended at version 2.7.1 in July 2025. TROPOMI is filtered using the main data quality flag retaining values greater than or equal to 75 (out of 100). No cloud filter is directly applied to TROPOMI, so all cloud filtering not included in the main quality flag is a result of TEMPO cloud filtering.

Sampling: Same as for NO2 described in section 3.2.2.

Metrics: Same as for NO2 described in section 3.2.2.

Aggregation: Same as for NO2 described in section 3.2.2.

Results

PSPI HCHO-01 requires TEMPO to be able to distinguish high values from background values. For PSPI HCHO-01, we look at the overall skill of TEMPO across the US and at individual sites. Figure 3-38 shows that TEMPO is able to capture some of the variation in formaldehyde across the US. Figure 3-39 shows TEMPO and TROPOMI at both Pandora and NAA regions. The NAA regions are spatially much larger than the Pandora regions, which explains why both TEMPO and TROPOMI HCHO variation is larger and more homogenous within the NAA regions. In many cases, a Pandora site is within a NAA region. The difference between the data at the Pandora and the NAA that contains it represents the urban/rural gradients within the NAA that are not present in the smaller Pandora region. For example, there are four Pandora stations (Cornwall-CT, Madison-CT, New Haven-CT, and Westport-CT) in the Connecticut NAA box. Both TROPOMI and TEMPO rank the sites in the same order from greatest to least: Westport, New Haven, Madison, Cornwall. Similarly, there are four Pandora stations (Philadelphia PA, Bristol PA, New Brunswick NJ, and Bayonne NJ) within the Philadelphia NAA box. For the Philadelphia NAA, TROPOMI and TEMPO do not agree on the rankings. This is an area of continuing research. Lastly, Figure 3-40 shows a strong correlation between TEMPO and TROPOMI (r=0.82) when comparing across geographical areas for Pandoras, but not for NAA regions (r=0.38). The individual sites large relative variation in both TEMPO and TROPOMI as shown by the error bars in Figure 3-40. The large relative variation highlights a challenge for formaldehyde that is not present in nitrogen dioxide.

The TEMPO validation plan requires an assessment of bias and precision for at least one month **(HCHO-02)**, two months in different seasons **(HCHO-04)**, and a year **(HCHO-06)**. This analysis reports several metrics of bias and correlation with respect to TROPOMI for over 2 years (August 2023 through July 2025). Figure 3-41 provides a temporal overview of the TEMPO and TROPOMI record. This comparison shows interquartile ranges (IQR: 25%-75%) and medians from each month to highlight sub-seasonal variability. Unlike tropNO2, the TotHCHO product has different temporal signatures in TEMPO and TROPOMI. The products agree quite well from May to early September, but TEMPO TotHCHO can be as little as half of TROPOMI in the winter when HCHO columns are lowest. The bias structure can be seen more clearly in Figure 3-42 where the biases and correlations (linear and orthogonal) are best May through August.

Figure 3-38, Figure 3-39, and Figure 3-40 show a strong spatial correlation between TEMPO and TROPOMI across sites. The dynamic range of observations by location can be seen in Figure 3-39. The normalized mean bias overall is -1.5×10^{15} molecules/cm^2 (-15%) overall and varies by sites. At Pandora sites, the mean bias average was -1.3×10^{15} (-12%) with a standard deviation of 1.1×10^{15} molecules/cm^2 (10%). At NAA regions, the mean bias average was -1.7×10^{15} molecules/cm^2 (-16%) with a standard deviation of 1.1×10^{15} molecules/cm^2 (10%). For the normalized mean bias at NAA regions, 23 sites are within 15%, 27 sites within 15-30%, and 1 site is above 30%. For normalized mean bias at Pandora locations, there are 28 sites within 15%, 13 sites between 15% and 30%, and 3 sites above 30%. The correlation by site also varies by NAA (Pandora) region with just 7 (17) sites having a correlation between 0.25 and 0.35 and the remaining 44 (21) sites having a correlation below 0.25. This highlights two things: 1) the skill varies by site and users should review the performance at their site; 2) the HCHO correlation at individual locations has less skill than NO2 at reproducing TROPOMI HCHO. Additional information by location including bias statistics and correlation values is available at https://gaftp.epa.gov/Air/aqmg/bhenders/share/TEMPO/ (available as of 2025-06-26).

In summary, the TEMPO TotHCHO product reproduces some of the spatial variability found in TROPOMI but has large temporal differences. TEMPO and TROPOMI disagree on the strength of the seasonal signal. They agree more closely in the summer, but TEMPO retrieves lower values than TROPOMI during winter. It is important to note that TROPOMI has a known high bias at low HCHO (Lambert et al., 2024) and should be viewed here as a correlative measurement rather than validation. The next section focuses on Pandora, which can more reasonably be considered a validation product.

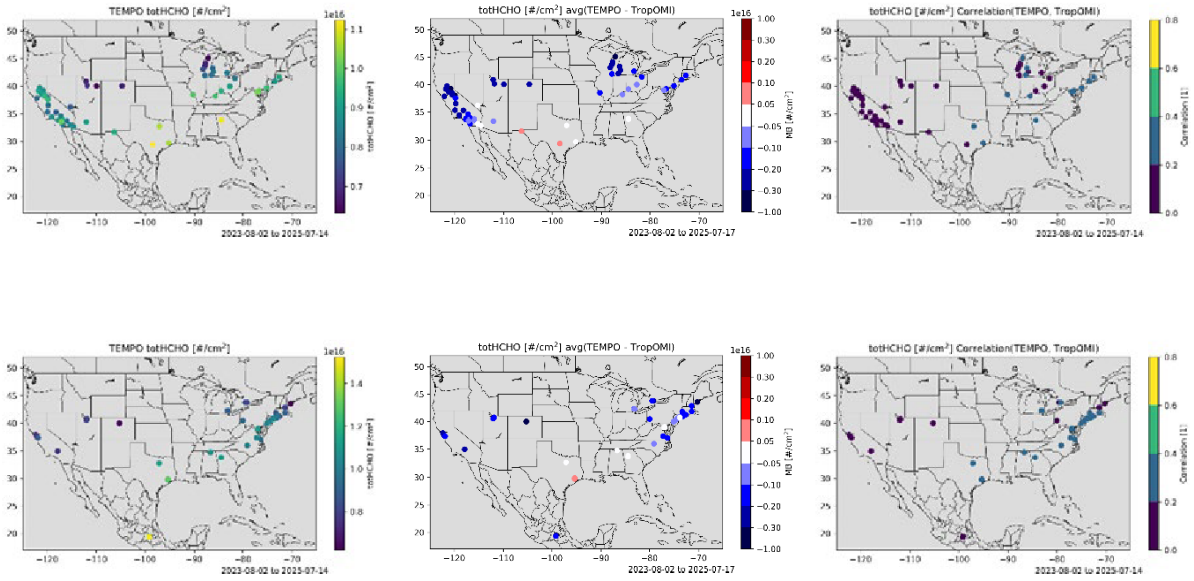


Figure 3-38: TEMPO (left), mean bias (middle) and correlation (right) TotHCHO at Ozone Non-Attainment Areas (top) and Pandora locations (bottom).

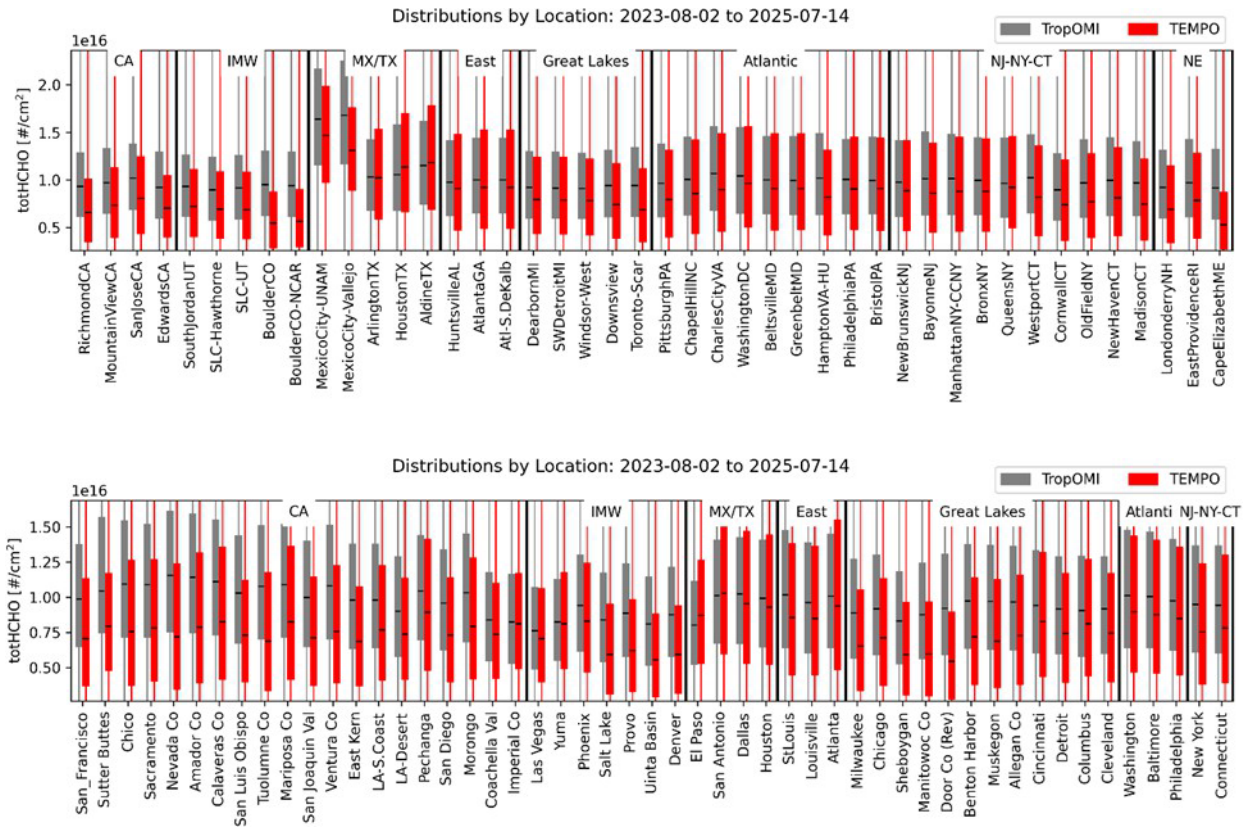


Figure 3-39 : TEMPO (red) and TROPOMI (grey) TotHCHO distributions at Pandora (top panel) locations and Ozone Non-Attainment Areas (bottom panel). The boxes represent the 25%, 50% and 75%, while the whiskers show the minimum and maximum. Locations within each panel are sorted into regions: California (CA), Intermountain West (IMW), Mexico/Texas (MX/TX), East, Great Lakes, Atlantic, and New Jersey-New York-Connecticut (NJ-NY-CT).

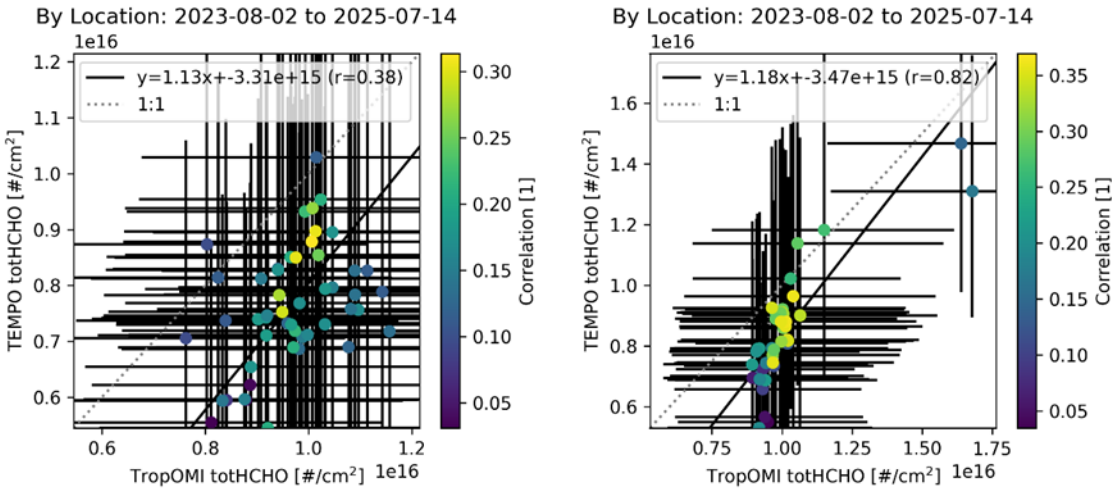


Figure 3-40: TEMPO (y-axis) vs TROPOMI OFFL (x-axis) TotHCHO shows strong correlation for Pandora locations (left) and more moderate for NAA (right) with lower values than TROPOMI throughout the distribution.

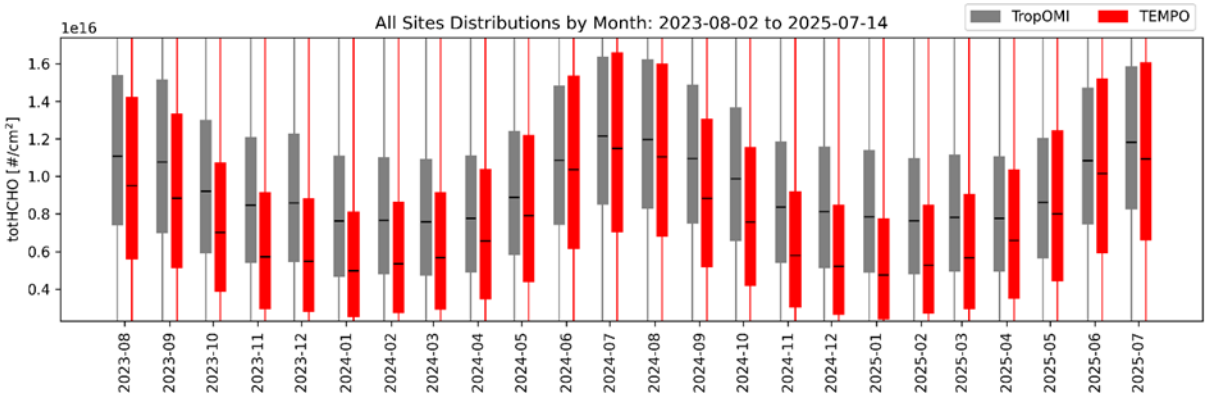


Figure 3-41: TEMPO (red) and TROPOMI OFFL TotHCHO interquartile ranges (line: median, lines: 25% to 75%) by month.

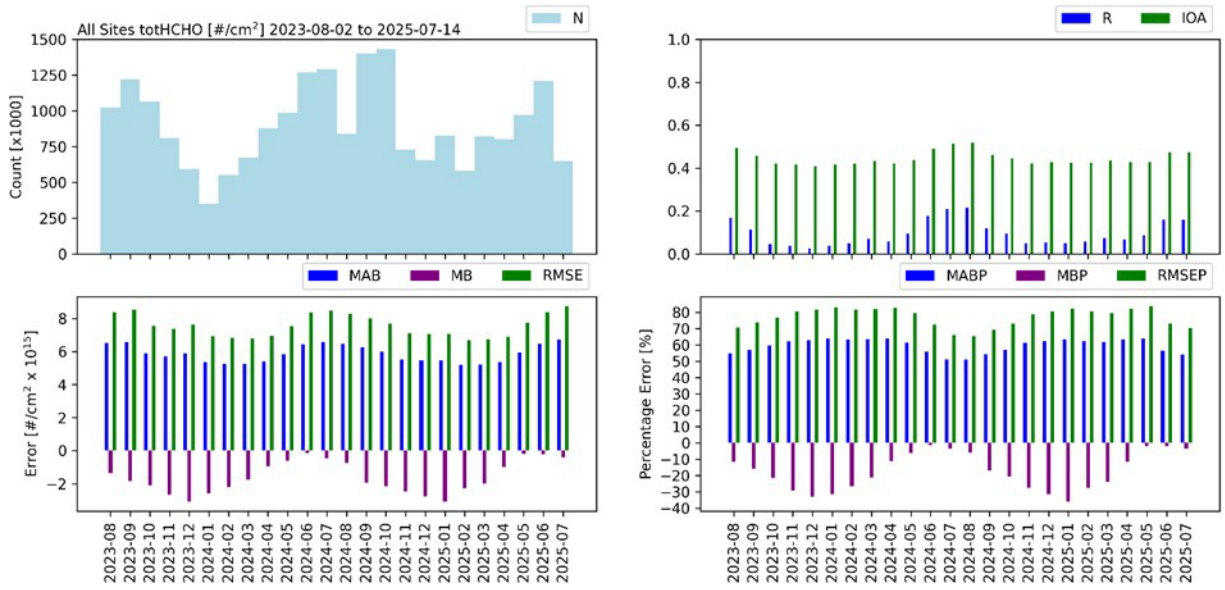


Figure 3-42: Statistics by month for TEMPO and TROPOMI TotHCHO. The top row shows the count (N), correlation (R), and index of agreement (IOA) by month. The bottom row shows monthly mean absolute bias (MAB), mean bias (MB), and root mean square error (RMSE) in molecules/cm2 on the left and as a percent of the mean on the right.

3.3.2 TotHCHO Inter-Comparison Between TEMPO and Pandonia Global Network

Authored by: Barron Henderson, Lukas Valin, Eric Baumann, Dave Williams, Todd Plessel and Jim Szykman (EPA); Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO-TEMPO)

Introduction

This section compares TEMPO to the ground-based Pandora HCHO products to address validation PSPIs HCHO-01, HCHO-02, and HCHO-04, and HCHO-06. The results here have also been used to guide analyses for PSPI HCHO-05.

Methodology

Datasets: TEMPO and Pandora total vertical column density (VCD) formaldehyde (HCHO) products are coincidently sampled at all Pandora sites within the TEMPO field of view. The TEMPO HCHO product that is the focus here is referred to as “version 3” that has data been reprocessed to cover the entire period of record (August 2023 to October 2024). The Pandora product discussed here is the “rfus5.1_8” version that is available through the Pandonia Global Network (PGN) and redistributed through the Remote Sensing Information Gateway. The Pandora HCHO is filtered for medium quality and above based on (Rawat et al., 2025). For formaldehyde, some stations with data are excluded due to data quantity or quality issues. Sites are removed for data quality when the direct sun measurement and sky scan (aka MAX-DOAS) measurement have low correlation (<0.25). Sites are removed for data quantity when there

are few measurements during the version 2 time-period. The following instruments by ID are removed: 31, 34, 51, 52, 72, 74, 139, 154, 181, 184, 185, and 204. As a result, we use the following instrument IDs: 25, 32, 38, 39, 57, 61, 64, 66, 69, 70, 80, 134, 135, 140, 142, 145, 147, 156, 157, 166, 179, 183, 186, 187, 207, 208, and 237. Updates to the geolocations of Atlanta instruments may affect performance in that region.

Sampling: Same as Pandora NO2 in Section 3.2.3

Metrics: Same as TROPOMI NO2 in Section 3.2.3

Results

Figure 3-43, Figure 3-44, and Figure 3-45 address aspects of the validation PSPI HCHO-01. Figure 3-43 shows TEMPO TotHCHO skill at individual sites across the field of view. TEMPO captures some of the broad variations across the TEMPO field of view. For example, the expected higher formaldehyde in the south is clearly evident. However, there are strong biases at some sites particularly in the northeast. For example, Figure 3-44 shows that at New Brunswick NJ and Manhattan NY the median TotHCHO from TEMPO is larger than Pandora's 75% percentile. Lastly, Figure 3-45 shows a good correlation (r=0.78) between TEMPO and Pandora when comparing across geographical areas. This highlights that TEMPO HCHO captures the variation across locations that is expected based on the reference product Pandora.

The TEMPO validation plan requires an assessment of bias and precision for at least one month (HCHO-02), two months in different seasons (HCHO-04), and a year (HCHO-06). This analysis reports several metrics of bias and correlation with respect to Pandora for nearly 2 years (Sept 2023 through July 2025). Figure 3-46 highlights sub-seasonal variability with monthly interquartile ranges (IQR: 25%-75%) and medians from both Pandora and TEMPO for all stations. TEMPO and Pandora agree both on the sub-seasonal variability as well as the longer-term seasonal variability. In contrast to TROPOMI HCHO, both TEMPO and Pandora show consistent decreases from summer into winter and increase back to summer. Figure 3-47 highlights the same consistency with accuracy and error statistics by month .

The performance at individual sites varies and is only summarized here. The mean bias (normalized mean bias) overall was -0.43e15 molecules/cm^2 (-3.9%) and varies by site. The individual site mean bias (normalized mean bias) had an average of 0.064e15 molecules/cm^2 (6.8%) and varies with a mean of -0.43e15 molecules/cm^2 (-3.9%) and a standard deviation of 2.1e15 molecules/cm^2 (24%). For the normalized mean bias 14 sites are within 15%, 8 sites are within 15-30%, and 6 sites above 30%. The correlation also varies by site with 12 sites having a correlation between 0.5 and 0.75, and 16 having a correlation less than 0.5. This highlights that the skill varies by site and users should review the performance at their site. Additional information by location including bias statistics and correlation values is available at https://gaftp.epa.gov/Air/aqmg/bhenders/share/TEMPO/ (available as of 2025-06-26).

Figure 3-48 and Figure 3-49 address the diurnal component of validation PSPI HCHO-02. Figure 3-48 shows that the interquartile range of TEMPO and Pandora HCHO evolve similarly over time. Both TEMPO and Pandora show lower HCHO in the morning that increases with time-of-day peaking in the late afternoon. Within each local solar time (LST=UTC-LON/15) hour, Figure 3-49 panels show the accuracy

and correlation between TEMPO and Pandora TotHCHO. Formaldehyde correlation and index of agreement are broadly consistent from 10 LST to 15 LST with marginally lower correlations from 6-10 LST and 15-18 LST. This pattern can also be seen in the low bias that is evident in the early morning and late evening.

In summary, the TEMPO TotHCHO performs reasonably well compared to Pandora. This is in contrast to the TROPOMI TotHCHO product where seasonal differences lead to larger differences. However, TROPOMI has a known high bias at low HCHO (Lambert et al., 2024). The Pandora measurement though not fully validated is the official validation product for TEMPO. As a result, we tend to weigh the results of the good performance against Pandora more heavily than TROPOMI. Although the Pandora performance is better, the performance of the TEMPO TotHCHO product is not as good as the tropNO2 or totNO2 products. This is expected due to the difficulty of HCHO retrieval.

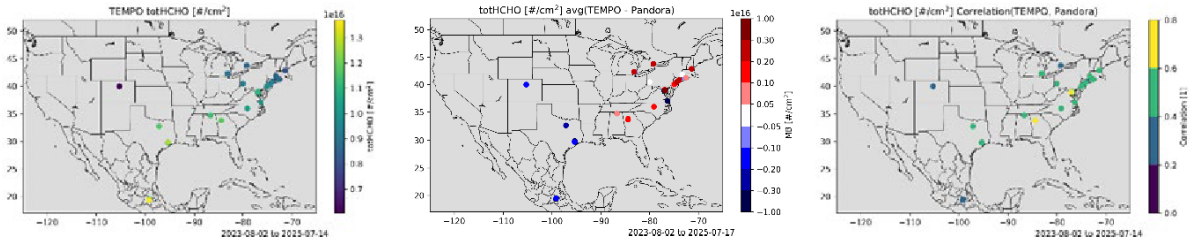


Figure 3-43: TEMPO (left), mean bias (middle) and correlation (right) in TotHCHO at Pandora locations.

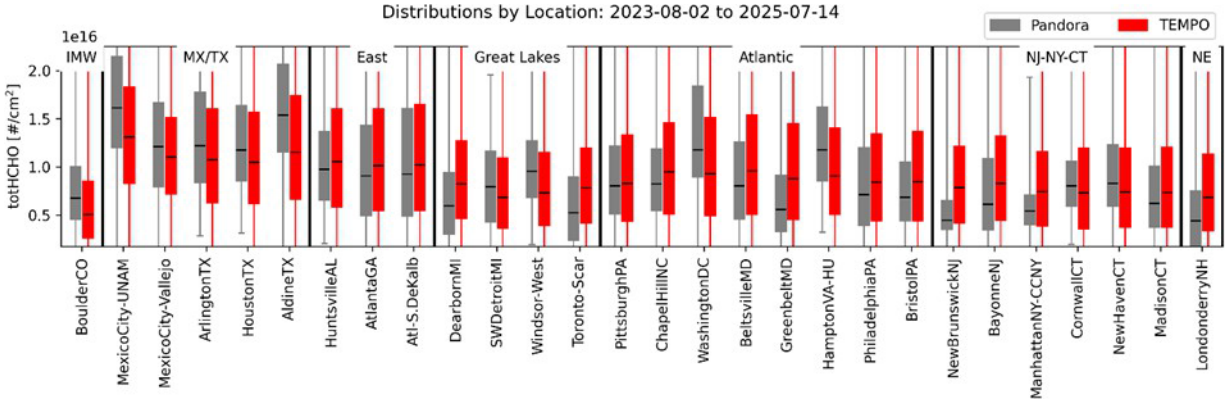


Figure 3-44: TEMPO (red) and Pandora (grey) TotHCHO distributions. The boxes represent the 25%, 50% and 75%, while the whiskers show the minimum and maximum. Locations within each panel are sorted into regions: Intermountain West (IMW), Mexico/Texas (MX/TX), East, Great Lakes, Atlantic, New Jersey-New York-Connecticut (NJ-NY-CT), and Northeast (NE).

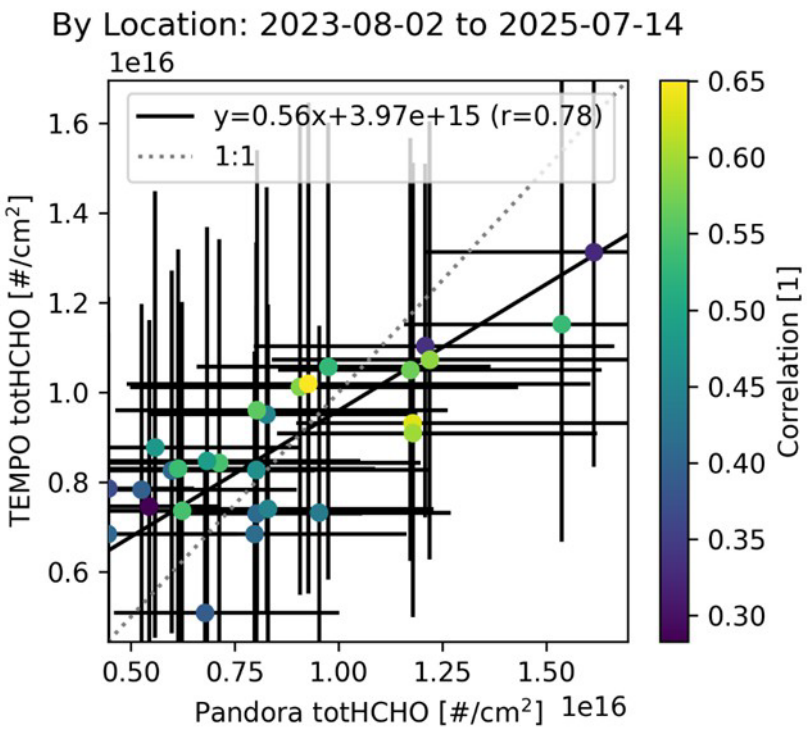


Figure 3-45: TEMPO (y-axis) vs Pandora (x-axis) in TotHCHO shows strong correlation with a low-bias indicated by the slope that is strongest at the high-end of the distribution.

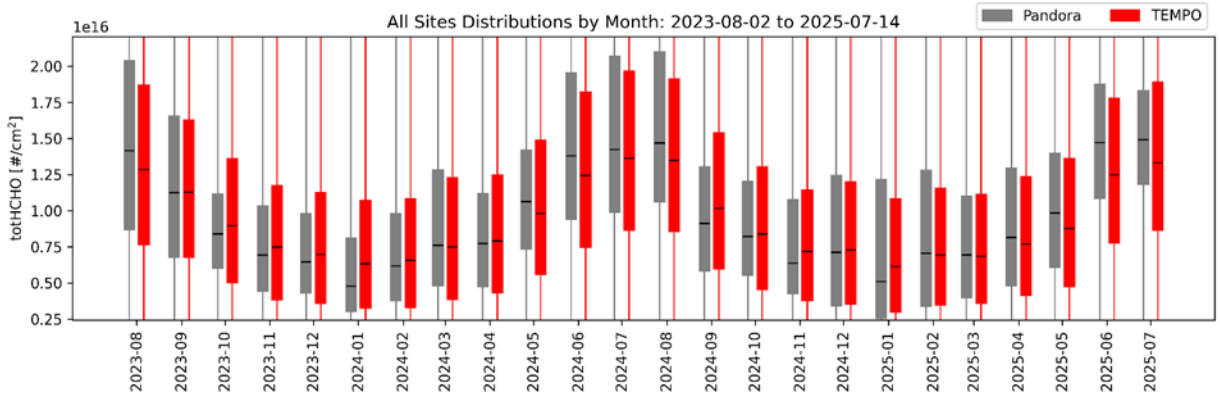


Figure 3-46: TEMPO (red) and Pandora TotHCHO interquartile ranges (marker: median, lines: 25% to 75%) for monthly intervals.

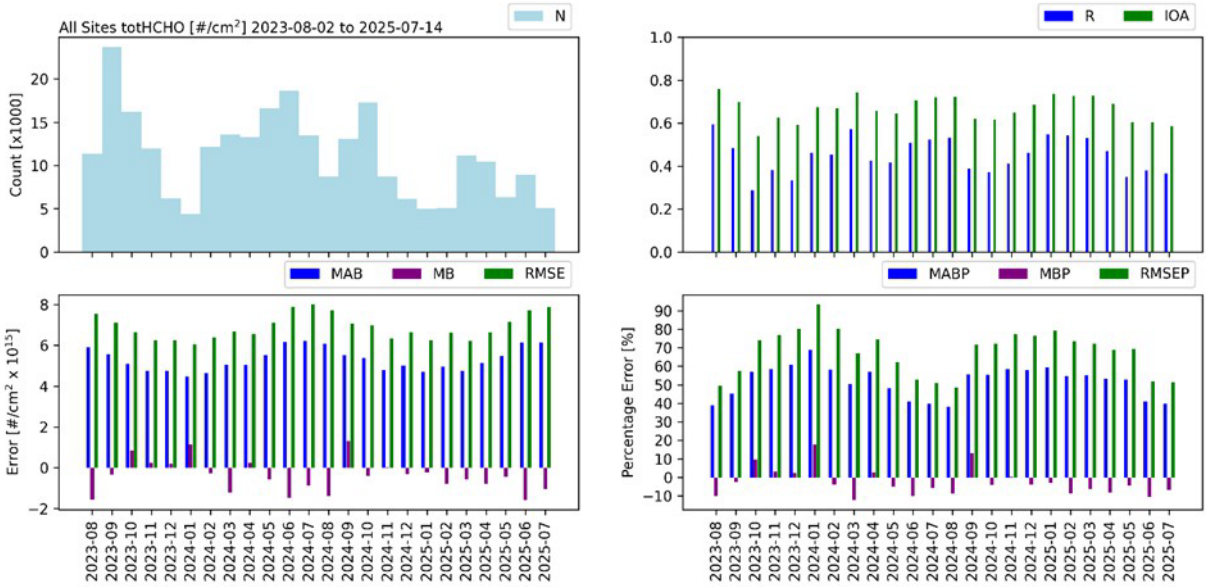


Figure 3-47: Statistics by month for TEMPO and TROPOMI TotHCHO with both linear regression and orthogonal regression best fit lines. The top row shows the count (N, upper left), correlation (R), and index of agreement (IOA) by month. The bottom row shows monthly mean absolute bias (MAB), mean bias (MB), and root mean square error (RMSE) in molecules/cm2 on the left and as a percent of the mean on the right.

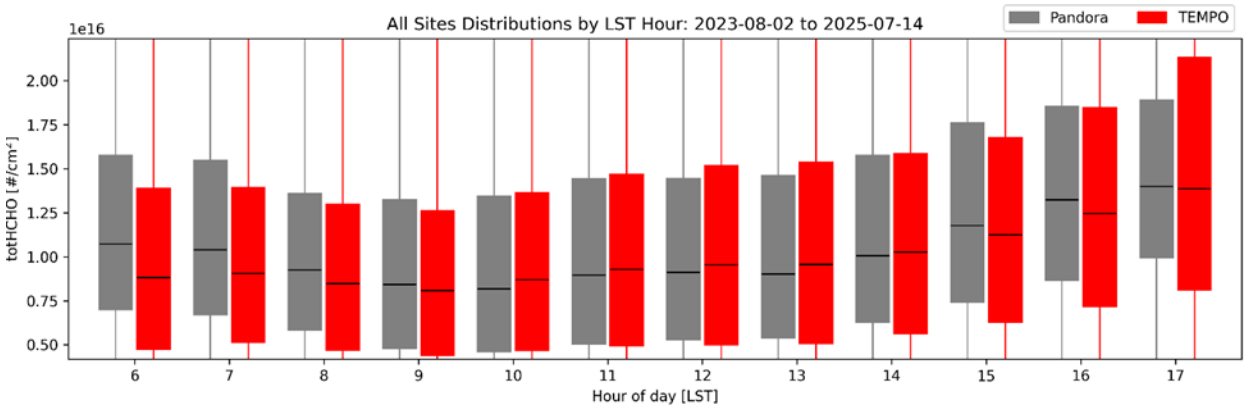


Figure 3-48: TEMPO and Pandora TotHCHO interquartile range as a function of local solar time (LST=UTC-LON/15) from 6LST to 17LST across all sites.

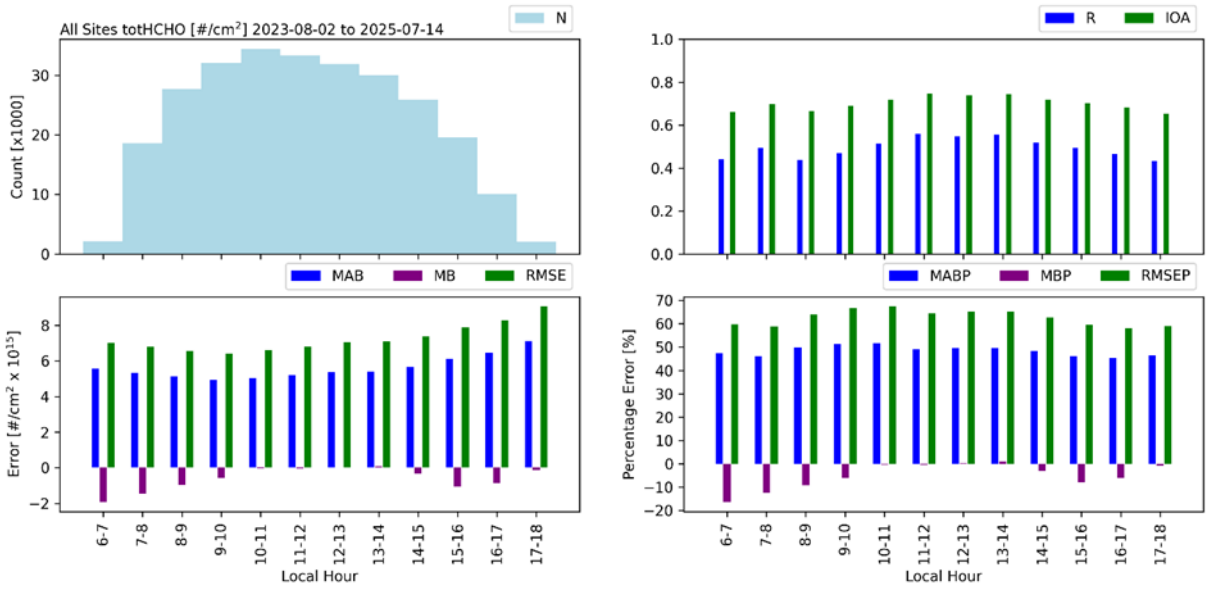


Figure 3-49: Statistics by local solar time (LST=UTC-lon/15) for TotHCHO from TEMPO (y-axis) and Pandora (x-axis) from 6LST to 17LST across all sites. The top row shows the count (N, upper left), correlation (R), and index of agreement (IOA) by month. The bottom row shows monthly mean absolute bias (MAB), mean bias (MB), and root mean square error (RMSE) in molecules/cm2 on the left and as a percent of the mean on the right.

3.3.3 TotHCHO Inter-Comparison Between TEMPO and Pandonia Global Network

Authored by: Prajjwal Rawat, Katherine Travis, Laura Judd, James Crawford, (NASA LaRC); Barron Henderson, Lukas Valin, Todd Plessel, , Eric Baumann, Jim Szykman (EPA); Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO-TEMPO)

Datasets: This analysis includes assessing the TEMPO HCHO "version 3" product between August 2023 and September 2024. TEMPO data are filtered using the following criteria: main_data_quality_flag > 0, eff_cloud_fraction > 0.2, and solar_zenith_angle > 70°.

As in the previous section, TEMPO HCHO will be evaluated with the Pandora product "rfus5.1_8" available through the Pandonia Global Network (PGN) and redistributed through the Remote Sensing Information Gateway. Pandora's performance for measuring HCHO columns can vary by instrument, with less well-performing instruments having artificial variations in HCHO columns. To address this issue, Pandora sites are chosen based on (1) Pandora direct-sun (DS) and sky-scan (SS) HCHO contemporaneous observations with correlation (R^2) greater than 0.35, (2) similar diurnal profile shapes between DS and SS, (3) sites having smaller biases between DS and SS HCHO column (< 50%), and (4) data completeness of more than 25% for the selected time period of August 2023-September 2024 between Pandora and TEMPO. The analysis below only uses Pandora DS HCHO columns. Thirty Pandora sites passed the selection criteria for DS HCHO. To maximize available Pandora DS HCHO data from these sites, we use the filtering method from Rawat et al. (2025) to ensure Pandora data quality.

Sampling: TEMPO data is sampled within a 10 km radius from each Pandora site and compared to Pandora data within a 10 min window from the TEMPO scan.

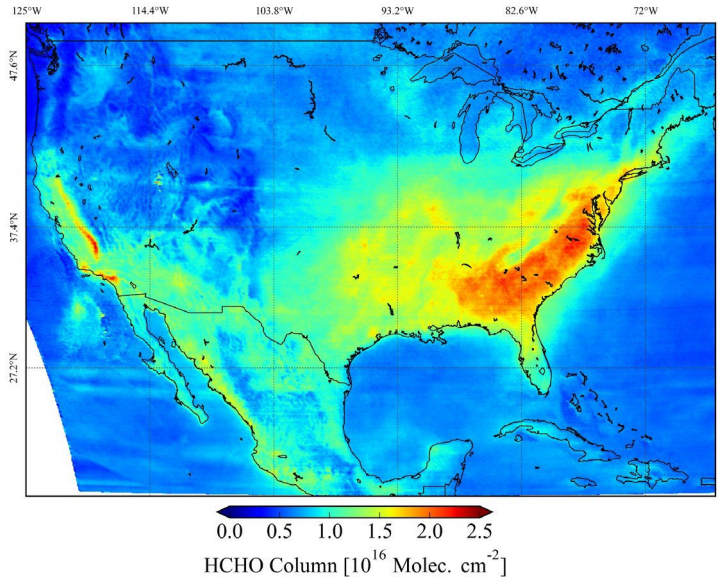
Metrics: Deming regression parameters are used to estimate the slope and intercept along with the coefficient of determination (R^2). The mean bias is calculated using the mean difference between TEMPO and Pandora [TEMPO - Pandora], while the mean percent bias is calculated using [(TEMPO-Pandora)/Pandora]*100. The root mean square error (RMSE) is also calculated using square root of mean squared error between TEMPO v3 and Pandora DS HCHO column.

Figure 3-50 (a) shows an average map of TEMPO HCHO columns over the CONUS for 10 days in July 2024 showing large scale regional gradients with the highest columns SE US and lowest values to the NW US and over water, as expected. Figure 3-50 (b) shows how TEMPO compares to data at the 30 Pandora sites selected across the field of regard. Overall, a good spatial correlation (R^2) of 0.60 was observed between TEMPO v3 and Pandora. The slope of 0.7 indicates that TEMPO underestimates HCHO in areas with larger HCHO columns but captures the variation across at these sites which span a range of locations. The mean absolute difference is calculated as $2.6x10^{15}$ molecules cm^{-2} with the majority of the spread well within the $10x10^{15}$ molecules cm^{-2} precision requirement for TEMPO in this averaged space.

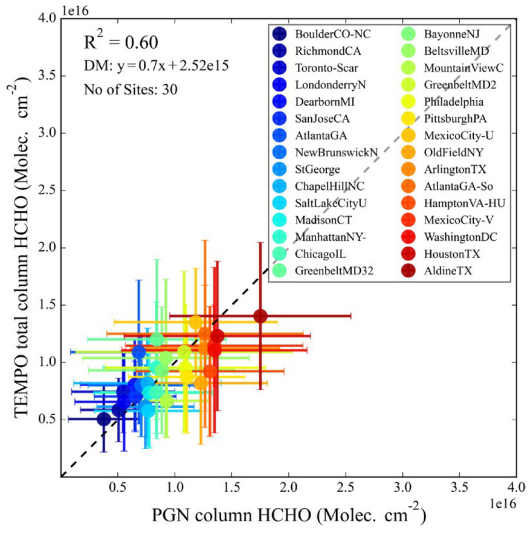
To further assess the performance of TEMPO across these sites, Figure 3-51 shows the statistics of the TEMPO v3 and Pandora DS HCHO intercomparison by site ordered by the mean Pandora HCHO column. The highest Pandora HCHO columns were observed in AldineTX and HoustonTX and the lowest columns were observed in BoulderCO. Twenty-four of the 30 sites have a mean % bias within ±25%. The bias approaches 60% only for GreenbeltMD32 and AtlantaGA which is due to the relatively lower Pandora DS column which may be a Pandora data quality issue (see discussion below). The bias is positive for sites with lower HCHO columns and trends negative as the HCHO column increases. R^2 spans between 0.1 and 0.8, with 19 sites having an R^2 above 0.5. RMSE is also well below the precision requirement for all sites with 20 sites below $5x10^{15}$ molecules cm^{-2}.

Figure 3-52 shows scatter plots of the data comparisons at each site colored by time of day to illustrate the dynamic range and variance of the comparison populations. Some sites show close to a 1:1 relationship, however, there are many sites that show TEMPO underestimating HCHO compared to Pandora at the highest values, aligning with the overall statistics by site (Figure 3-50). Figure 3-53 breaks this analysis apart by time of day with box plots representing the range of data each hour by site. Diurnal patterns agree well at most sites with R^2 often exceeding 0.5 by hour. However, each site does deserve its own in-depth examination when analysis is being applied by region. A few sites worthy of noting here as not being in agreement diurnally. For example, there are two locations in Greenbelt MD; one site agrees quite well (PGN #2) whereas the other instrument greatly disagrees with TEMPO (PGN #32). This points toward an issue with Pandora data quality despite the level of evaluation done on the Pandora datasets prior to use in this analysis. Another location set is AtlantaGA vs. AtlantaGA-So. This analysis provides a glimpse into the results by site, but more study is still needed to investigate diurnal and seasonal biases at individual sites, including continued scrutiny of the Pandora data quality.

(a)



(b)



(c)

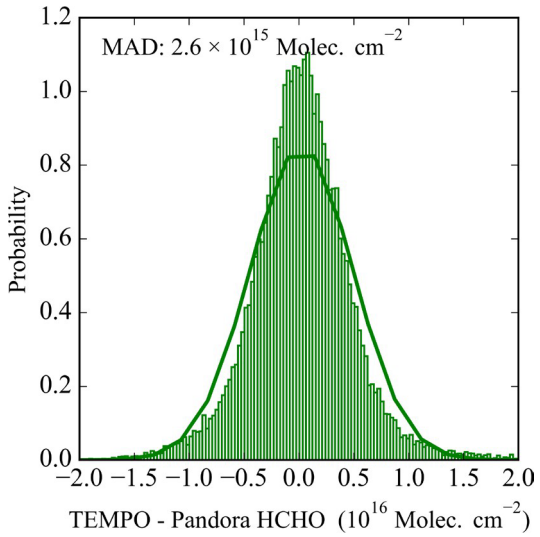


Figure 3-50 (a) Spatial map of TEMPO L3 HCHO column for 10 days in July 2024 using level-2 data oversampled at 1x1 km^2. (b) Spatial correlation of HCHO column between TEMPO v3 and Pandora direct sun HCHO column over the 30 selected Pandora sites in the PGN during August 2023-September 2024. (c) Histograms showing the difference between TEMPO and Pandora HCHO columns with a fitted Gaussian distribution shown by solid curve line with annotated Median Absolute Deviation (MAD).

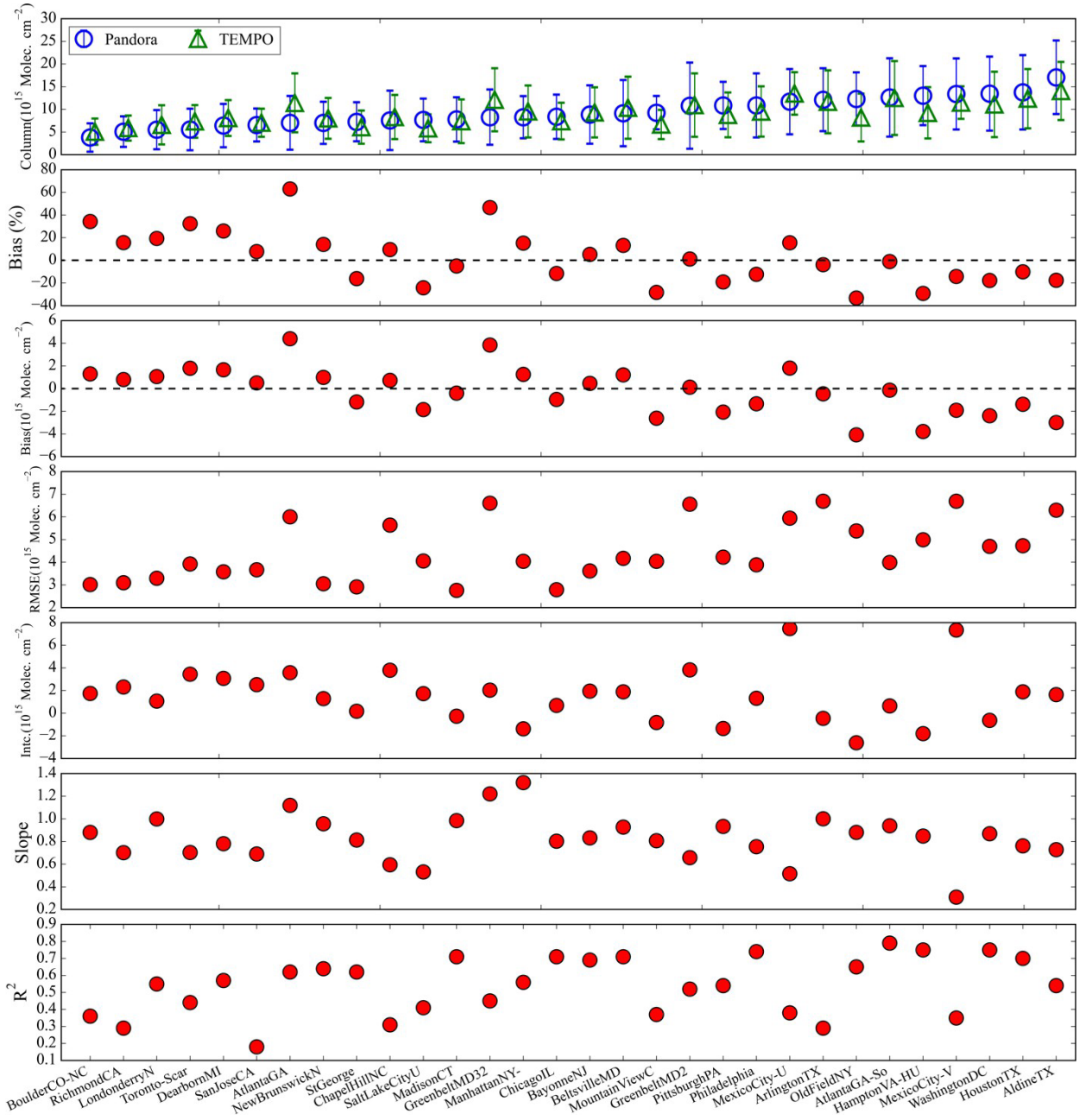


Figure 3-51 Statistical analysis between the Pandora and TEMPOv3 HCHO column over individual Pandora sites between Aug 2023 to Sep 2024. In the top panel the mean column with standard deviation for each selected Pandora site is shown from Pandora (blue) and TEMPO (green) in ascending order of Pandora HCHO column.

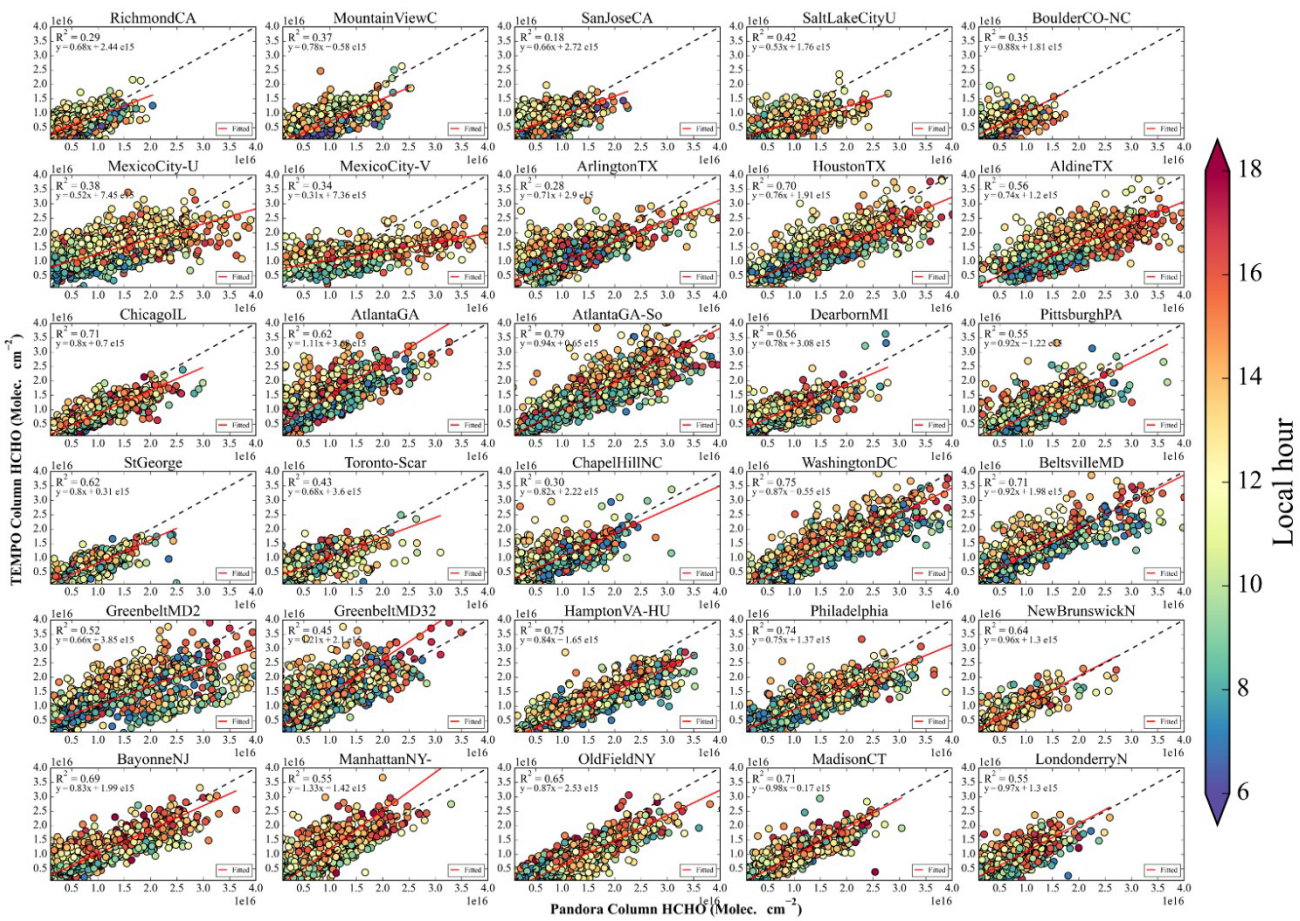


Figure 3-52 Scatter plot between Pandora and TEMPO v3 HCHO column over selected Pandora sites between Aug 2023-Sep 2024 colored by time of day.

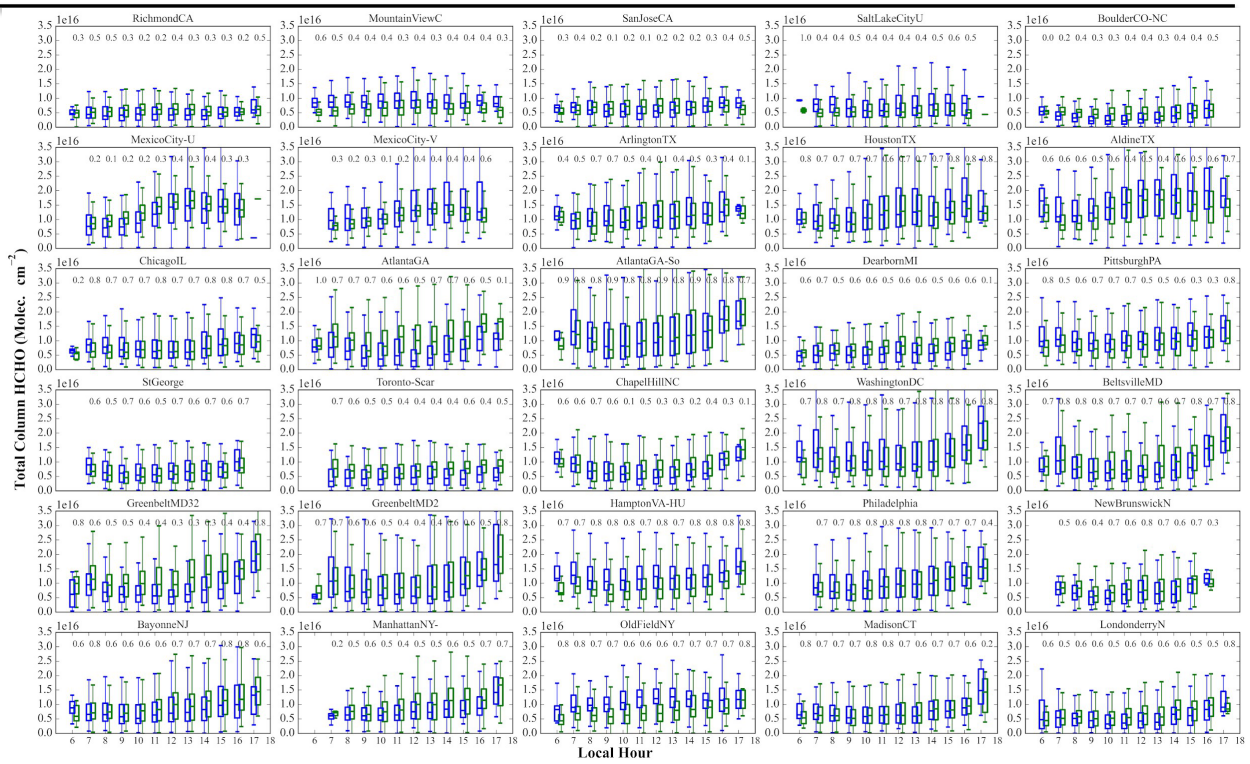


Figure 3-53 Diurnal variation of HCHO column from TEMPOv3 and Pandora direct sun measurements during Aug 2023 to Sep 2024. The Pandora sites are arranged from west to east from top left to bottom right. TEMPO (green) and Pandora (blue) observations are shown with a box plot next to each other. The R2 of different hour are also annotated at the top of each hour.

3.3.4 TotHCHO Inter-Comparison Between TEMPO and the Network for the Detection of Atmospheric Composition Change (NDACC) Fourier transform infrared (FTIR)

Authored by: Ivan Ortega, Sara Martinez-Alonso, Jim Hannigan, David Edwards (ACOM/NCAR); Wolfgang Stremme, Andrea Cadena, Michel Grutter (UNAM); Kimberly Strong, Victoria Flood (U. Toronto); Xiaoyi Zhao (ECCC); Gonzalo Gonzalez Abad, Caroline Nowlan (SAO-TEMPO)

In this effort, we evaluate the quality of the Level 2 (L2) version 3 (V3) HCHO TEMPO product using ground-based solar-absorption FTIR (Fourier Transform InfraRed) measurements from three stations within the TEMPO field of regard: Mexico City, Mexico; Boulder, Colorado; and Toronto Canada. The ground-based observations follow the strategies recommended by the Infrared Working Group (IRWG) within the Network for the Detection of Atmospheric Composition Change (NDACC; https://ndacc.larc.nasa.gov/).

Table 3-4 lists information about the ground-based FTIR stations included in this analysis. These stations perform regular solar absorption measurements under clear-sky conditions. For HCHO analysis, they use retrieval settings optimized and harmonized across all IRWG/NDACC stations (see Vigouroux et al. (2018) for retrieval details). The uncertainty budget for each site, as listed in Vigouroux et al. (2020), ranges between 5-8% for random errors and approximately 12% for systematic errors. The degrees of freedom for signal (DOFS) is around 1.1 at all sites, indicating that only vertical columns are provided without vertical profile information.

Table 3-4: FTIR stations that contribute to the present work.

Station	Latitude	Longitude	Altitude (km)	Instrument	Code	Team
Mexico City	19.33°N	99.18°W	2.26	Bruker Vertex 80 (0.004 cm^{-1})	PROFFIT	UNAM
Boulder	40.04°N	105.24°W	1.61	Bruker 120/5 HR (0.004 cm^{-1})	SFIT4	NCAR
Toronto	43.60°N	79.36°W	0.17	Bomem DA8 (0.06 cm^{-1})	SFIT4	University of Toronto

The analysis presented here covers the period from August 2023 to June 2024, though data for early 2024 is incomplete.

To compare with TEMPO, we averaged several pixels to reduce the random uncertainty of the TEMPO mean HCHO data and increase sensitivity to day-to-day variability. We tested radii of 2, 5, 10, 15, and 20 km around FTIR sites. High variability was observed in TEMPO data using a 2 km radius, but starting from 5 km, this variability diminished. For this analysis, we used mean values within a 10 km radius of the sites. Data was filtered using the recommended quality criteria: main quality flag different from 0, solar zenith angle larger than 70°, and cloud fraction larger than 0.4. If more than 50% of the pixels were flagged, the mean values were not used.

Figure 3-54 illustrates an example of HCHO total columns retrieved by TEMPO for a single scan over North America (a) on August 11, 2023. The panels b-d display the three sites where we have the FTIR measurements: Mexico City, Boulder, and Toronto, each marked with a black circle representing a 10 km radius around the extraction points. The mean number of TEMPO pixels within this radius are 40 for Mexico City, 27 for Boulder, and 25 for Toronto.

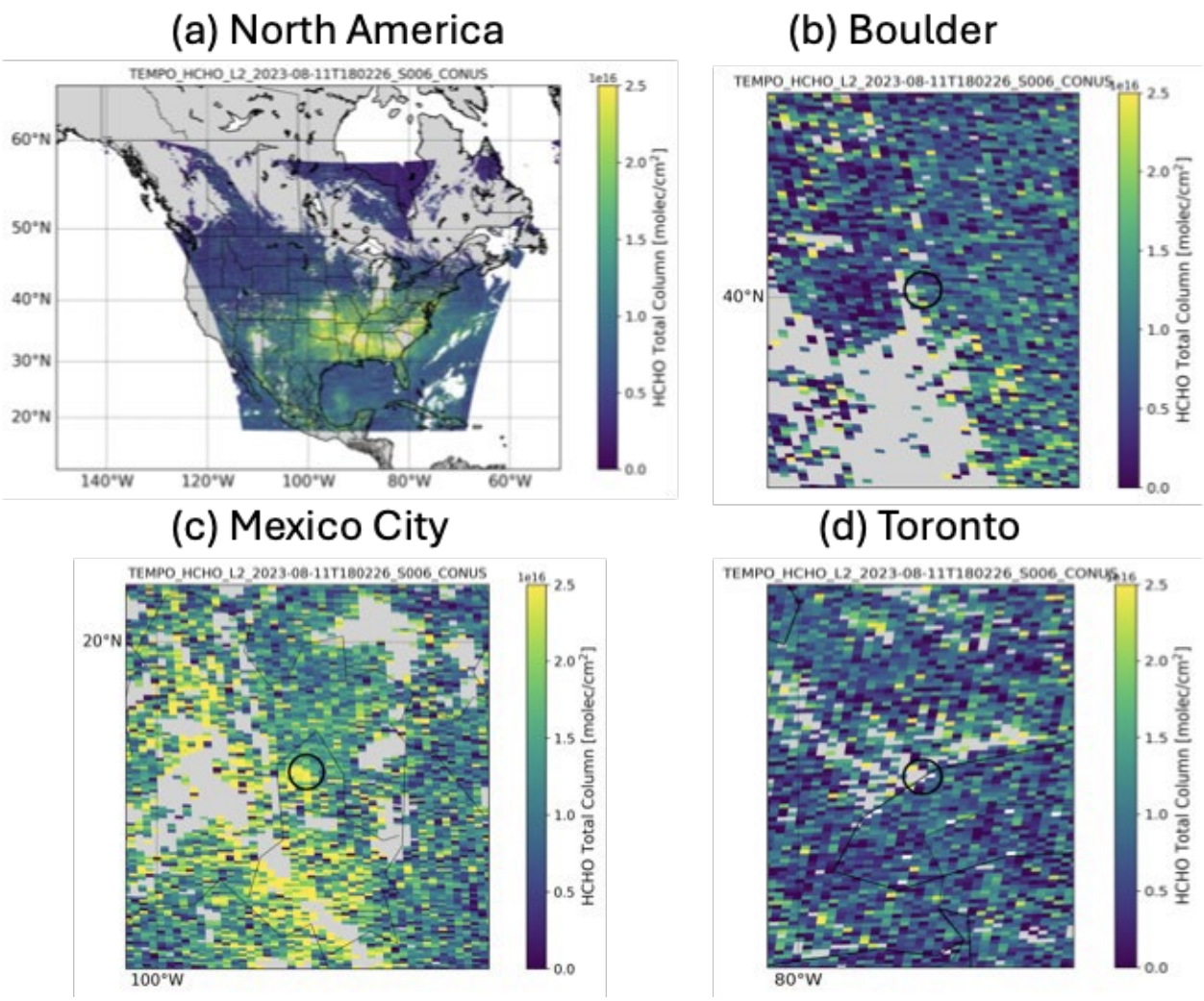


Figure 3-54: Example of the TEMPO HCHO total column for a single day on August 11, 2024, at 18:00 UT. The upper left panel displays North America (a), while panels b, c, and d show Boulder, Mexico City, and Toronto, respectively. The black circles around each site indicate a 10 km radius for TEMPO data extraction.

The analyzed time series are illustrated in Figure 3-55. TEMPO total columns are displayed in blue, and FTIR total columns are shown in gray. The green points represent the FTIR total columns at the time of the TEMPO overpass. These values are generated for each day when a TEMPO overpass occurs, with the FTIR data interpolated to match the exact time of the TEMPO overpass. For each day we avoid interpolation when there is a gap of 2.5h or larger in ground-based observations.

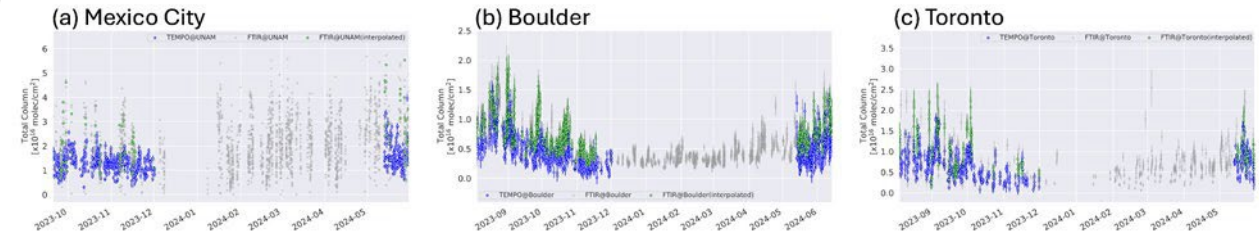


Figure 3-55: The time series of HCHO total columns at the three sites are shown. TEMPO total columns are displayed in blue, while FTIR total columns are shown in gray. The FTIR data interpolated to the exact time of the TEMPO overpass are represented in green. Note that TEMPO data is missing for some months in 2023 and early 2024.

Figure 3-56 shows correlation plots for all coincident points. Table 3-5 summarizes the percent biases at each station. We applied Orthogonal Distance Regression analysis, with the slope, intercept, and R-value displayed for each site. The bias of the TEMPO HCHO measurements is estimated by calculating the median of the relative differences between the collocated TEMPO and reference FTIR data at each station:

$$BBBBAASS = mmggmm\left(\frac{TTEEAATTTT_{ii} - AATTBBRR_{ii}}{AATTBBRR_{ii}}\right) \qquad EEEE.\ 3.3.4.1$$

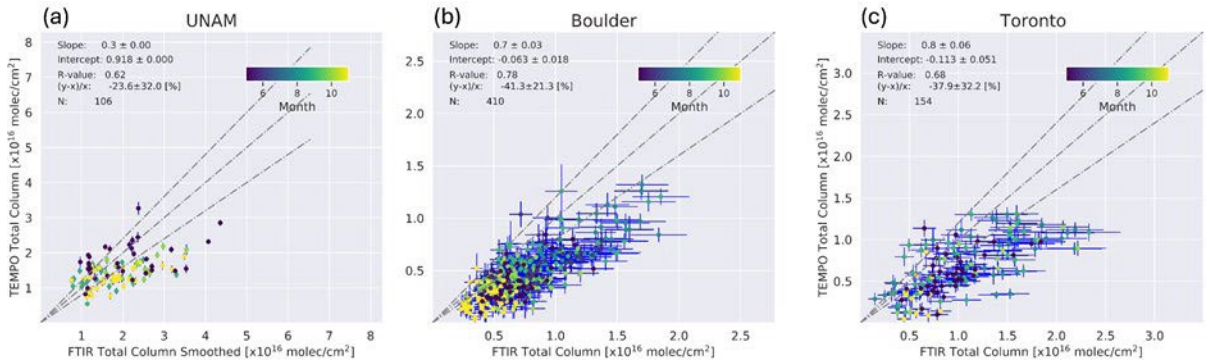


Figure 3-56: Scatter plots of TEMPO versus FTIR for individual collocated pairs color coded by month. Results of the linear Orthogonal Distance Regression are shown in each panel.

Table 3-5: Summary with median Bias estimated using equation 3.3.4.1.

Station	Bias [TEMPO-FTIR %]	slope	intercept (x10^{16} molec/cm^2)	r^2-value
Mexico City	-31.0 ± 38.8	0.3	0.90	0.38
Boulder	-41.3 ± 21.3	0.7	-0.06	0.61
Toronto	-37.9 ± 32.2	0.8	-0.11	0.46

3.3.5 TotHCHO Inter-Comparison Between TEMPO and In situ profiles from NASA DC-8 from AEROMMA

Authored by: Eleanor Waxman, Andrew Rollins (NOAA-CSL), Glenn Wolfe, Abby Sebol (NASA-Goddard)

The comparisons to TEMPOS formaldehyde (HCHO) product, this analysis uses observations from the NASA Goddard In-Situ Airborne Formaldehyde (ISAF) instrument (Cazorla et al. 2015). ISAF detects HCHO directly via Laser Induced Fluorescence (LIF) at 353 nm. For AEROMMA, the accuracy is 15% and precision is 50 pptv at 1 Hz.

A similar analysis as outlined in Section 3.2.4 was done for formaldehyde (HCHO) data from the DC-8 in comparison to TEMPO. The primary differences between the NO2 and HCHO analyses are that 1) the TEMPO HCHO retrieval does not require a temperature correction and 2) there is no stratospheric/tropospheric separation so only total slant columns and total vertical columns are retrieved. Therefore, the relevant equations are:

$$SSVVVV_{ttttttHHHHHHTT} = \sum_{zz=0}^{ttttttttooaattaattaattheettee} nn_{iiii\ aaiittss}(zz) \times \Delta zz \times SS(zz) \qquad (EEEE.\ 3.3.5.1)$$

$$VVVVVV_{ttttttIIHHHHHHTT} = \sum_{zz=0}^{ttttttttooaattaattaattheettee} nn_{iiii\ aaiittss}(zz) \times \Delta zz. \qquad (EEEE.\ 3.3.5.2)$$

$$VVVVVV_{ttttttHHHHHHTT} = \frac{SSVVVV_{ttttttHHHHHHT\ \ T}}{AAAAAA_{ttttttTTTTTTTTT}} \qquad (EEEE.\ 3.3.5.3)$$

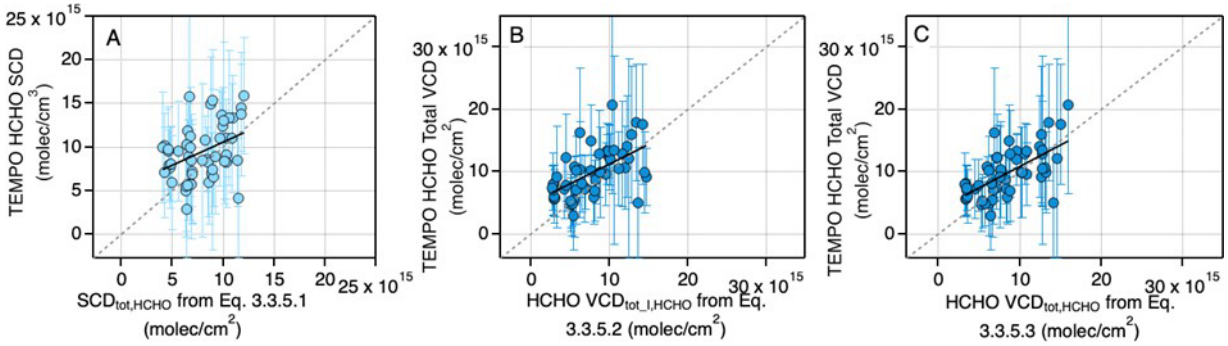


Figure 3-57: Scatter plots of TEMPO vs. DC-8 in-situ derived parameters for HCHO: (A) TEMPO SCD vs. in-situ derived SCD from DC-8 data using Eq. 3.3.5.1, (B) TEMPO Total VCD vs independent in-situ derived VCD from DC-8 data and TEMPO SW using Eq. 3.3.5.2, (C) TEMPO Total VCD vs in-situ derived VCD from DC-8 data including TEMPO priors using Eq. 3.3.5.3.

Table 3-6: Summary Statistics for TEMPO DC8 HCHO Comparisons

Panel	Compared quantity	Slope (unitless)	Intercept (molec/cm^2)	R^2	RMSE (molec/cm^2)	RMSE (%)
A	Slant columns	0.53	5.21×10^{15}	0.15	1.49×10^{15}	26.1%

B	Total vertical columns	0.63	$4.97x10^{15}$	0.32	$2.03x10^{15}$	43.2%
C	Total vertical columns	0.70	$3.9x10^{15}$	0.40	$1.31x10^{15}$	26.1%

Figure 3-57 shows the collection of intercomparisons between the NASA DC-8 derived parameters for HCHO and TEMPO. Summary statistics are provided in Table 3-6. In comparison to NO_2, the correlation is much lower (ranging from 0.15-0.40) as there is more scatter in the HCHO. The increased scatter between adjacent TEMPO pixels is also evident in the substantially larger error bars for HCHO relative to NO_2. However, there remains a clear relationship between the DC-8 derived and TEMPO HCHO products with similar RME percentages as for the NO_2 analysis.

3.3.6 TotHCHO Evaluation of TEMPO evaluation using WRF-Chem

Authored by: R. Bradley Pierce, Jerrold Acdan, Maggie Bruckner (UW-SSEC) Laura Judd, Scott Janz (NASA), Gonzalo Gonzalez Abad, Caroline Nowlan, Xiong Liu, Kelly Chance (SAO)

We use the same indirect validation approach outlined for NO2 in Section 3.2.7 to evaluate TEMPO V3 HCHO using WRF-Chem. However, for HCHO both GCAS and TEMPO are smoothed with a box-car smoother. GCAS HCHO has been preliminarily compared to Pandora HCHO direct-sun and sky scan measurements for sites in the STAQS domain that meet the criteria for site selection discussed in Section 3.3.3. Mean bias in GCAS data ranges from $-5x10^{14}$ for sky-scan vs $-45x10^{14}$ molecules cm^{-2} for direct sun with RMSE ranging from $33\text{-}63x10^{14}$ molecules cm^{-2}.

Figure 3-58 shows scatter plots of the WRF-Chem predictions and the STAQS GCAS column HCHO for each of the three urban areas (Chicago, Toronto, and New York City) sampled during the Eastern Phase of the NASA STAQS campaign (July 26 through August 16, 2023). The WRF-Chem HCHO columns are reasonably well correlated with GCAS within each urban area (r=0.31 to 0.54) with biases ranging from $25.3x10^{14}$ +/- $42.x10^{14}$ to $73.2x10^{14}$ +/- $43.x10^{14}$ molecules cm^{-2} and RMSE ranging from $54.6x10^{14}$ to $91.7x10^{14}$ molecules cm^{-2}. These comparisons establish the overall biases and RMSE between the validation data (remotely sensed HCHO) and the WRF-Chem CTM.

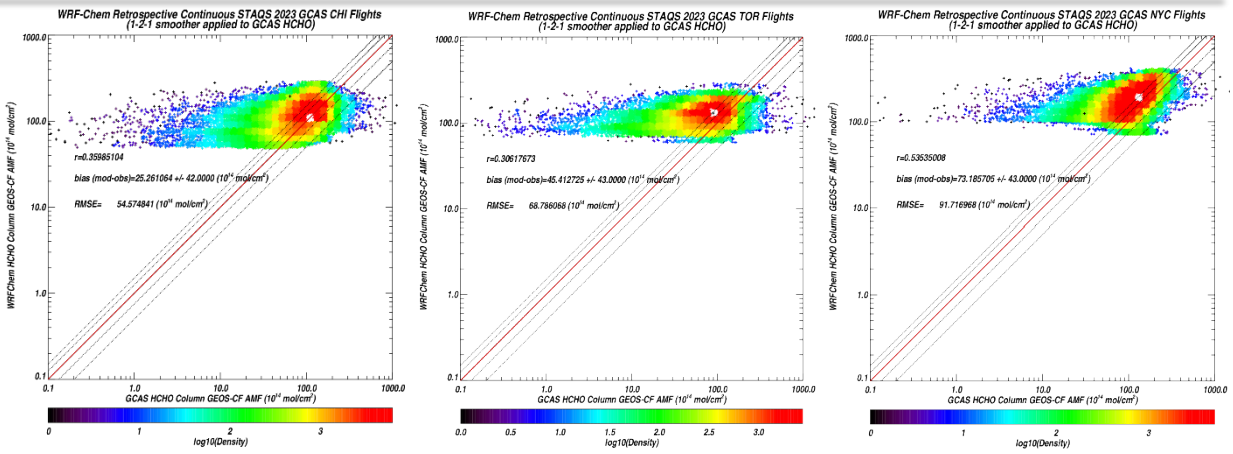


Figure 3-58: Scatter plots of WRF-Chem verses GCAS column HCHO for Chicago (CHI, left), Toronto (TOR, middle) and New York City (NYC, right). GCAS column HCHO was provided by L. Judd. Solid line in the scatter plot is the 1-to-1 line, dashed lines show 25% and 50% errors.

Figure 3-59 shows scatter plots of the WRF-Chem predictions and the TEMPO V3 box-car averaged column HCHO within the regional boxes defined in section 3.2.7 for all TEMPO observations over each of the cities sampled during the Eastern Phase of AEROMMA/STAQS. The WRF-Chem HCHO columns are reasonably well correlated with TEMPO within each urban area (r=0.52 to 0.64) with biases ranging from $17.1x10^{14}$ +/- $27.x10^{14}$ to $38x10^{14}$ +/- $33.x10^{14}$ molecules $cm^{-2}$and RMSE ranging from $32x10^{14}$ to $57x10^{14}$ molecules cm^{-2}.

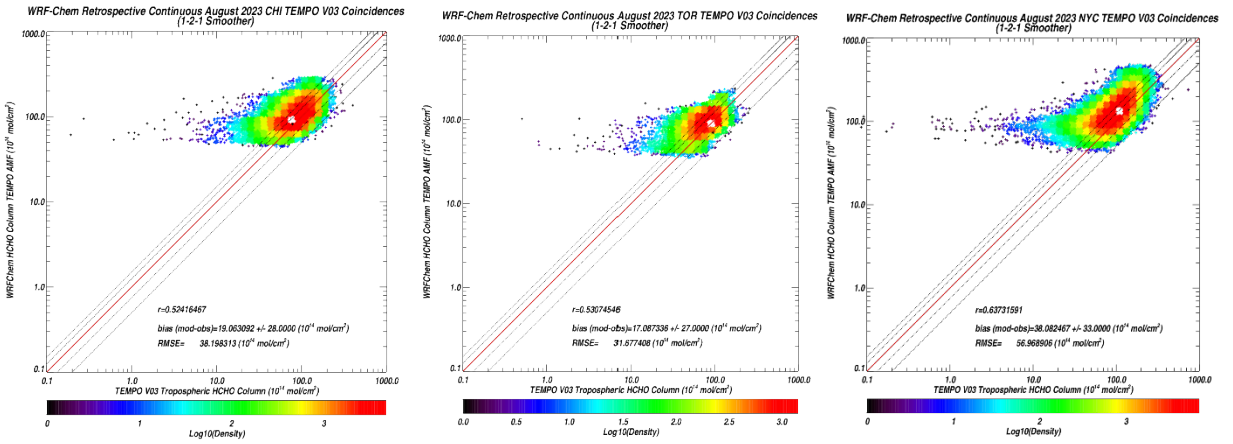


Figure 3-59: Scatter plots of WRF-Chem verses TEMPO V3 column HCHO for Chicago (CHI, left), Toronto (TOR, middle) and New York City (NYC, right). Solid red line in the scatter plot is the 1-to-1 line, dashed lines show 25% and 50% errors.

By taking the difference between the WRF-Chem - GCAS (Figure 3-58) and WRF-Chem - TEMPO (Figure 3-59) biases we can infer the TEMPO biases relative to GCAS. This results in inferred TEMPO-GCAS biases of $6.2x10^{14}$ +/- $50.5x10^{14}$, $28.3x10^{14}$ +/- $50.8x10^{14}$, and $35.1x10^{14}$ +/- $54.2x10^{14}$ molecules cm^{-2} for Chicago, Toronto, and New York City, respectively. The RMSE between WRF-Chem and TEMPO are lower than the RMSE between WRF-Chem and GCAS and are significantly less than the TEMPO HCHO column

precision requirement of $100x10^{14}$ molecules cm^{-2}. Some of this RMSE comes from uncertainties in the WRF-Chem simulations and so this is likely an overestimate of the true TEMPO RMSE.

To evaluate high/low and diurnal variations in the TEMPO HCHO column retrievals we sub-set the WRF-Chem simulations and TEMPO V3 HCHO retrievals into “southern (Eastern VA) ” and “northern (Lake Ontario) ” samples.

The sub-domains are defined as:

Southern (Eastern VA) Domain: 37N<Latitude<39N, 78.5W<Longitude<76.5W

Northern (Lake Ontario) Domain: 43N<Latitude<45N, 78.5W<Longitude<76.5W

Figure 3-60 shows the results of this evaluation. The TEMPO HCHO column retrieval is able to distinguish between high (Eastern VA) and low (Lake Ontario) regions, where the mean HCHO columns are $140x10^{14}$ and $70x10^{14}$ molecules cm^{-2}, respectively. The high (Eastern VA) region shows a stronger diurnal variation in the TEMPO HCHO column than the low (Lake Ontario) region. The WRF-Chem simulated HCHO column (with the TEMPO airmass factor and scattering weights applied) and the GEOS-CF a priori are both higher than TEMPO over the southern (Eastern VA) region while only the WRF-Chem HCHO column is higher than TEMPO over the northern (Lake Ontario) region.

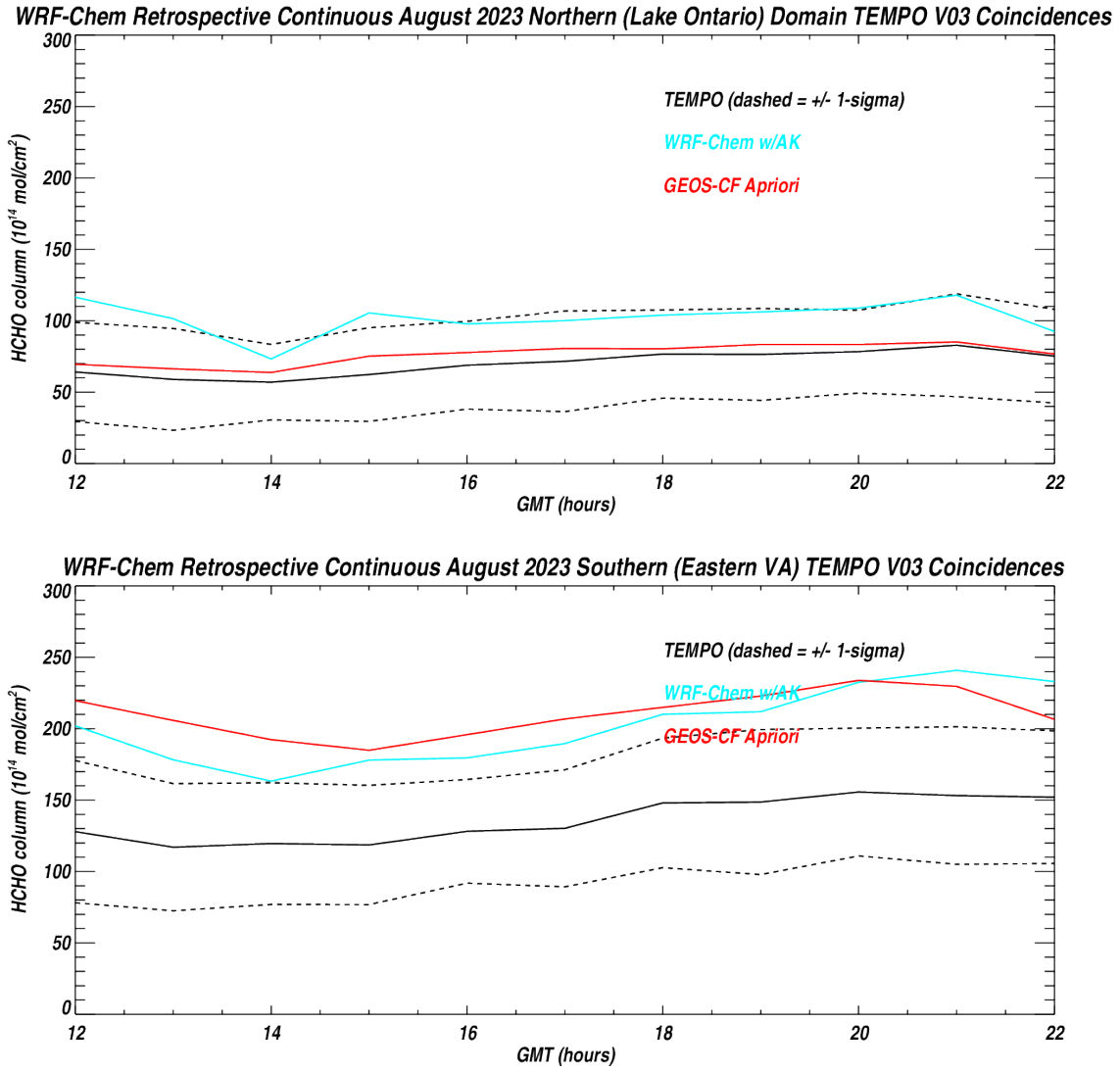


Figure 3-60: Mean diurnal variations in TEMPO (black), WRF-Chem (blue), and the GEOS-CF a priori (red) HCHO columns during August 2023 for Northern (Lake Ontario) Domain(upper), and Southern (Eastern VA) Domain (lower). The dashed black lines show +/- one standard deviation (1-sigma) for the TEMPO HCHO.

3.4 Level 2 Science Product Validation and Evaluation TEMPO_O3 L2_V03 - TEMPO O3 total column version 03

3.4.1 TotO3 Inter-Comparison Between TEMPO with Polar-Orbiting Satellite Nadir-Mappers

Authored by: J. R. Ziemke, S. M. Frith, N. A. Kramarova (NASA GSFC), Kanghyun Baek (Pusan University)

TEMPO V3 total ozone was tested against total ozone from OMI, OMPS, and TROPOMI satellite nadir-mappers. The time periods for TEMPO data included for the evaluations were December 2023 and May 2024. TEMPO total ozone maps (at ~6-minute increments) were matched up with OMI, OMPS, and TROPOMI daily measurements for these comparisons.

There have been no changes made to the TEMPO L2 algorithm going from V2 to V3, but L1 data input to V3 included a stray light correction. This stray light correction resulted in improvements below the 320 nm wavelength range. The analysis here indicates a latitude-dependent bias is present in TEMPO V3 total ozone and is similar to the previous TEMPO V2 total ozone bias.

OMI and OMPS Comparisons.

Figure 3-61 shows difference maps of TEMPO minus OMPS and TEMPO minus OMI total ozone averaged over several days in December 2023 (a-b) and May 2024 (c-d). There is a clear and persistent latitude-dependent offset where TEMPO V3 ozone is too large in lower latitudes and too small in higher latitudes.

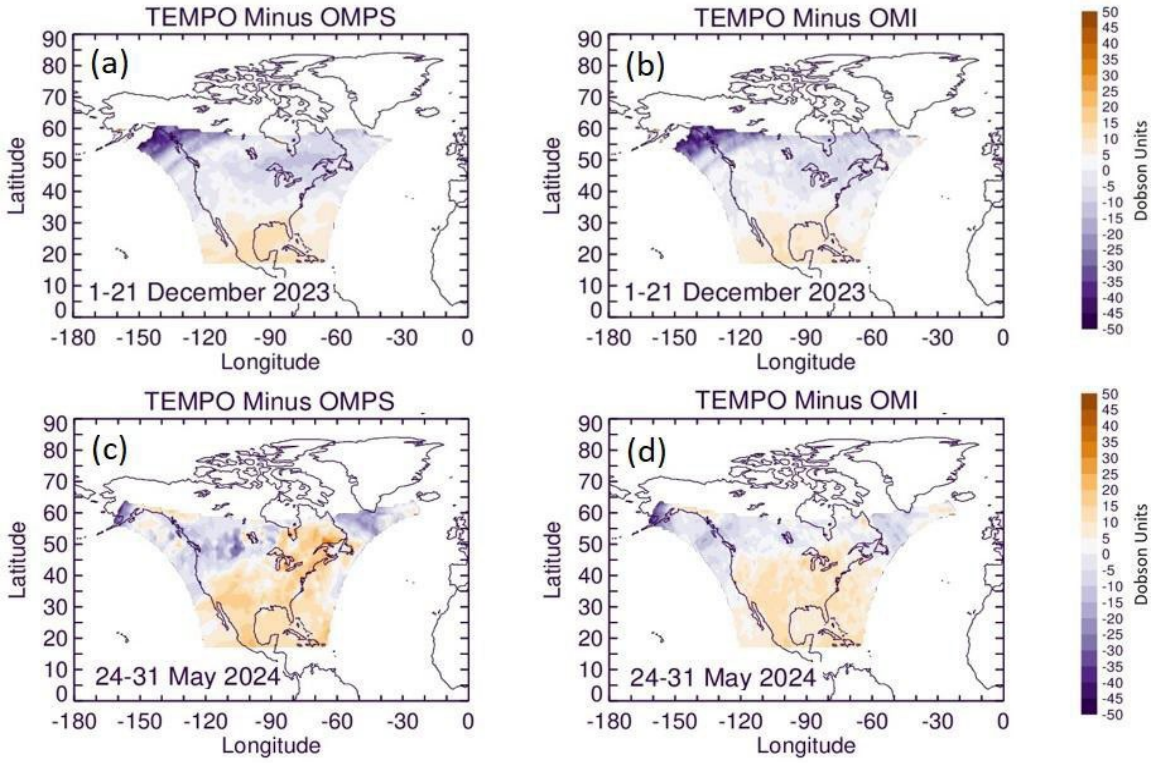


Figure 3-61: (a) Average differences (in Dobson Units) of TEMPO v3 minus OMPS total ozone for 1-21 December 2023. (b) Same as (a), but for TEMPO minus OMI. All data were binned for daily averages at 1o × 1o binning. Only TEMPO measurements lying within ±2 hours from either OMI or OMPS overpass were included in comparisons. (c-d) Same as (a-b) but for 24-31 May 2024.

Figure 3-62 shows offset differences for TEMPO v3 total ozone by averaging the data in Figure 3.4.1.1 along longitude and expressed in percent. Compared to both OMPS and OMI, the TEMPO total ozone is too high by ~3-5% in low latitudes and too low by several percent in high latitudes.

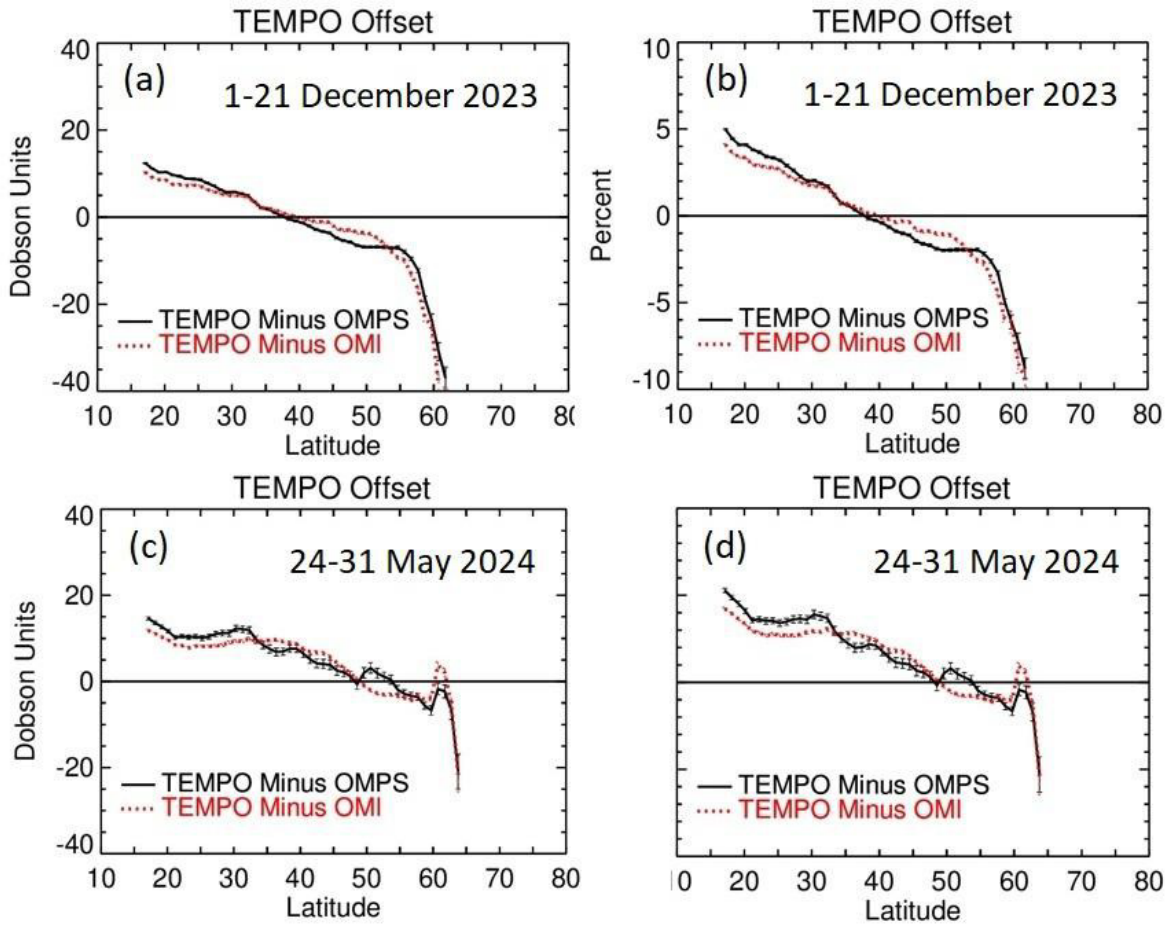


Figure 3-62: (a) Zonal-mean differences (in Dobson Units) of TEMPO v3 minus OMPS (black) and OMI (red) total ozone averaged along longitude for 1-21 December 2023. (b) Same as left panel, but instead plotted as percent relative to average total ozone (separately from both OMPS and OMI). Included are also standard error bars of mean differences which are very small. (c-d) Same as (a-b) but for 24-31 May 2024.

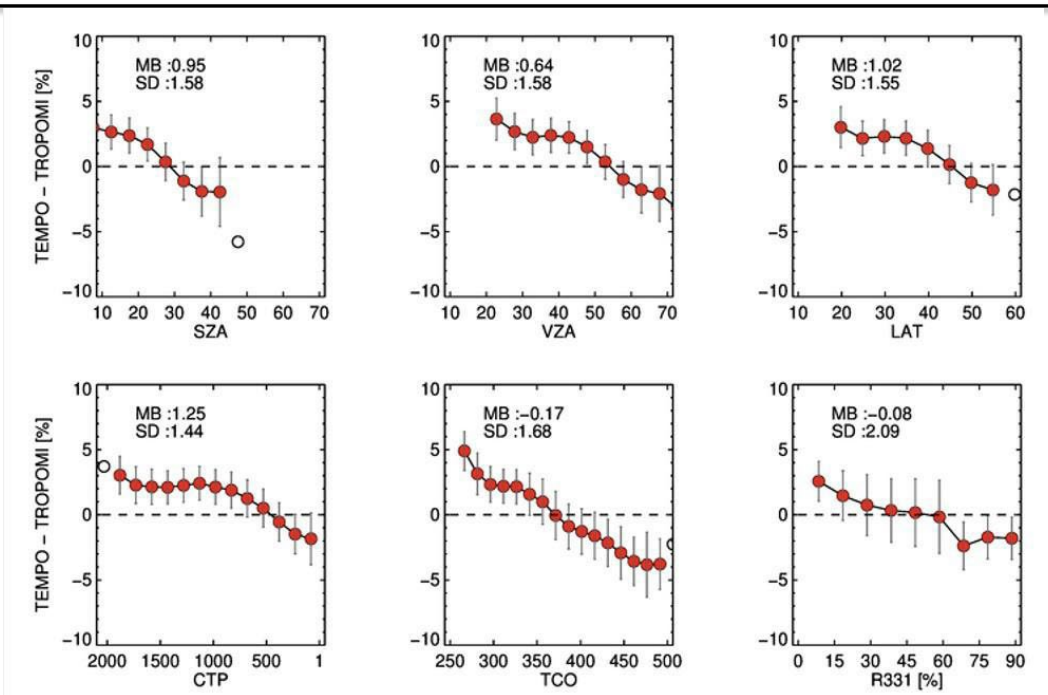


Figure 3-63: Percent differences of TEMPO Minus OMPS total ozone for 15-30 May 2024 as a function of solar-zenith angle (SZA), viewing zenith angle (VZA), latitude, cloud-top pressure (CTP), total column ozone (TCO), and 331 nm Reflectivity (R331) (indicated along the horizontal axes). Included are standard deviation bars. Only G03 granule data (at approximately 16:00 UTC) for TEMPO were included. All TEMPO data were re-gridded to 0.1o × 0.1o binning for comparisons. Geometry conditions for TEMPO for these comparisons are SZA < 45o and VZA < 65o. Coincidence criteria: distance < 10 km, time difference < 60 m, and CF < 0.8.

Figure 3-63 shows a more detailed analysis of TEMPO minus OMPS differences as a function of several different retrieval parameters (indicated along the horizontal axes). In all cases, the TEMPO V3 total ozone relative to OMPS total ozone is too high by 3-5 % for lower solar and viewing zenith angles, latitude, total column ozone, and 331 nm reflectivity and is too high by 3-5% for higher cloud top pressures.

TROPOMI Comparisons.

Figure 3-64 (left panel) shows TEMPO V3 total ozone minus TROPOMI total ozone differences for 15-30 May 2024 using only TEMPO G03 granule data. Figure 4 (right panel) shows a scatter diagram of the differences plotted in the left panel. The TEMPO offset differences relative to TROPOMI varies from +10 to +15 DU in low latitudes to -10 to -15 DU in high latitudes.

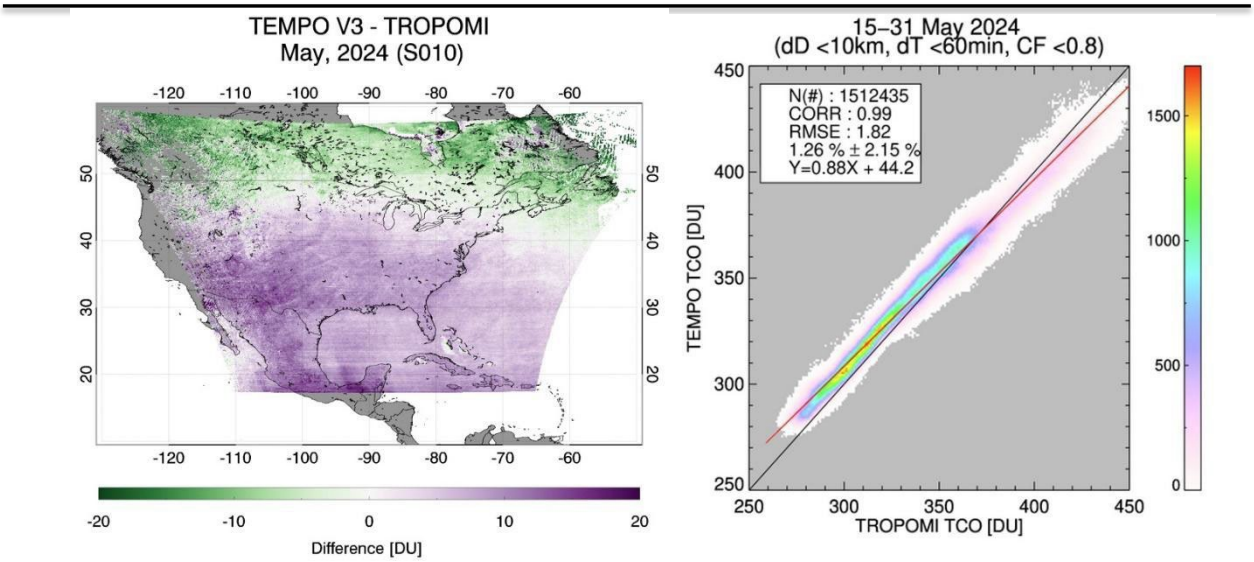


Figure 3-64: (left) Average difference of TEMPO V3 minus TROPOMI total ozone for 15-30 May 2024. (Right) Scatter plot of the TEMPO V3 minus TROPOMI total ozone differences plotted in the left panel. All TEMPO data were re-gridded to 0.1o × 0.1o binning for comparisons. Geometry conditions for TEMPO for these comparisons are SZA < 45o and VZA < 65o. Coincidence criteria: distance < 10 km, time difference < 60 m, and CF < 0.8.

Figure 3-65 shows analysis of TEMPO minus TROPOMI total ozone. The six panels in Figure 3-65 are the same aa in Figure 3.4.1.2 but with TROPOMI replacing OMPS. A systematic offset in Figure 3-65 is again present for all six panels. The offsets for TEMPO relative to TROPOMI are smaller than in Figure 3-62 with numbers of +2 to +3 % in low latitudes.

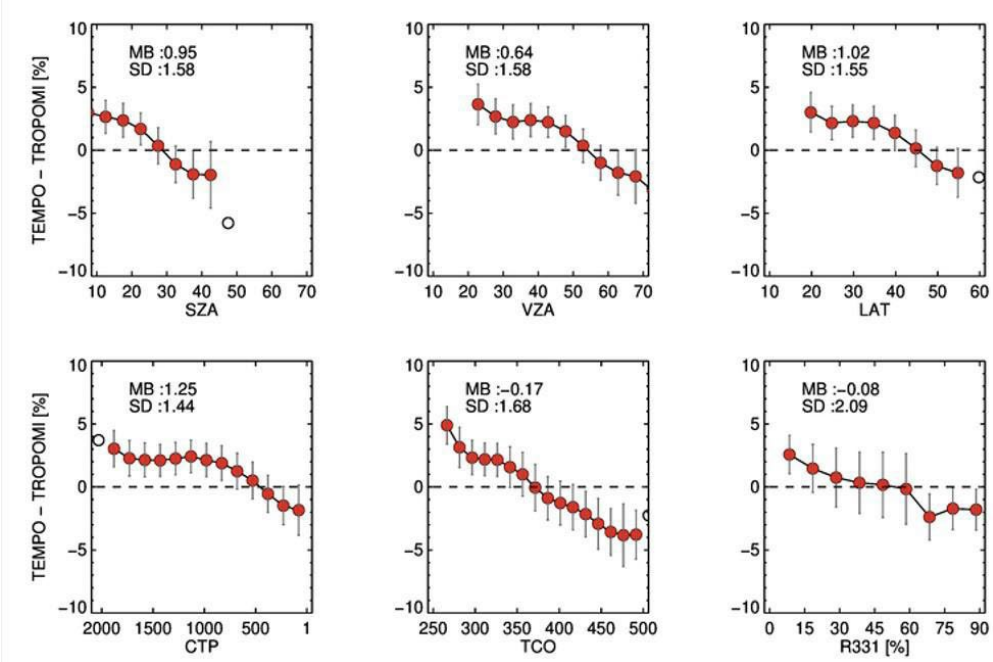


Figure 3-65: Percent differences of TEMPO Minus TROPOMI total ozone for 15-30 May 2024 as a function of solar-zenith angle (SZA), viewing zenith angle (VZA), latitude, cloud-top pressure (CTP), total column ozone (TCO), and 331 nm Reflectivity (R331) (indicated along the horizontal axes). Included are standard deviation bars. Only G03 granule data (at approximately 16:00 UTC) for TEMPO were included. All TEMPO data were re-gridded to 0.1o × 0.1o binning for comparisons. Geometry conditions for TEMPO for these comparisons are SZA < 45o and VZA < 65o. Coincidence criteria: distance < 10 km, time difference < 60 m, and CF < 0.8.

3.4.2 TEMPO ozone validation via ground-based network measurements

Authored by: Xiaoyi Zhao, Vitali Fioletov, Debora Griffin, Chris McLinden, Sum Chi Lee (Environment and Climate Change Canada), Irina Petropavlovskikh, (National Oceanic and Atmospheric Administration), Thomas Hanisco, (NASA GSFC), Lukas Valin, Eric Baumann, James Szykman (U.S. EPA). Alexander Cede, Martin Tiefengraber, Manuel Gebetsberger (LuftBlick)

TEMPO total column ozone (TCO) v3 data has been compared with ground-based measurements from Brewer and Dobson spectrophotometers and Pandora spectrometers. Total ozone measurements from two Dobson sites, four Brewer sites, and 58 Pandora sites were included in the analysis (see Fig. 3-66a). When compared with ground-based networks, TEMPO shows a stronger latitude dependence than TROPOMI (see Figs. 3-66b and c). For Pandonia Global Network, one outlier has been identified (AldineTX) and removed from the following analysis. Field calibration is needed to correct the bias observed at this site. Current official Pandora total ozone data (processing version rout2) are using information from climatology (TOMSv8) to estimate the ozone effective temperature. As a result, a weak positive latitude dependency is observed when compared with TROPOMI (Fig. 3-66c). More Brewer and Dobson observations from the U.S. will be included in future to further examine this issue.

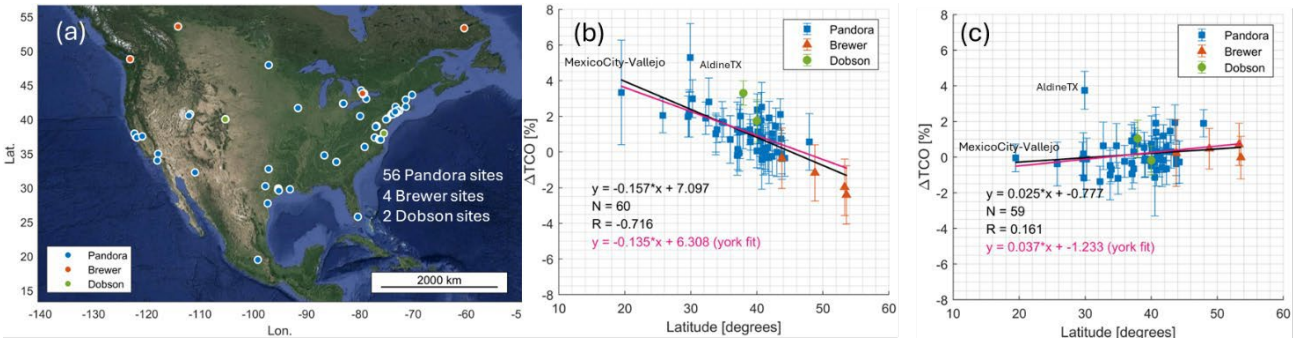


Figure 3-66: (a) Site map of Brewer, Dobson, and Pandora total ozone observations from 62 sites that have been included to perform validation for TEMPO V3 total column ozone. (b) Percentage difference between TEMPO and ground-based observations vs latitudes of the sites. (c) same as (b) but using TROPOMI total ozone observations.

In Fig. 3-67, the data has been further binned by 5 degrees latitude bins, to understand TEMPO's latitude dependency. In general, the multiplicative bias between TEMPO and ground-based measurements is very low (0.5%). For the six latitude bins, the range of the multiplicative bias is from -1.3% to 3.5%, with the <25 degrees bins showing the highest bias (which only has one Pandora site, MexicoCity-Vallejo).

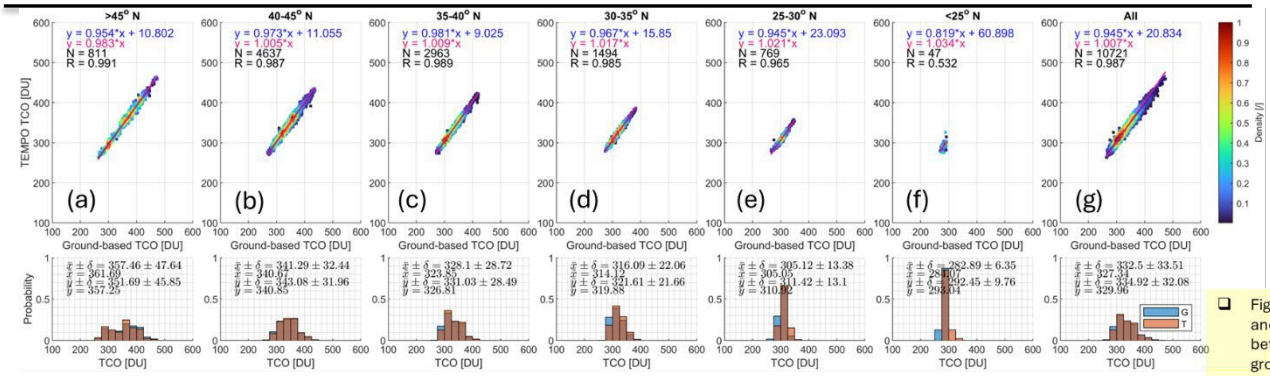


Figure 2. Scatter plots and histograms between TEMPO and ground-based measurements.

Figure 3-67: Scatter plots and histograms between TEMPO and ground-based measurements for different 5-degree latitude bins.

When compared to ground-based measurements, which have filtered out observations with slant ozone > 1000 DU, TEMPO TCO data shows an SZA dependency (see top row of Fig. 3-68). Simple empirical correction has been made to correct this SZA dependency (for SZA ≤ 70°).

$$TTVVTT_{cctttttteecctteecc} = TTVVTT_{tteeaatttt} \times (5.29gg^{-4}SSSSAA + 0.97) \qquad \text{Eq. 3.4.2.1}$$

The correction parameters were acquired using ground-based network observations. The corrected results are shown in the second row of Fig. 3-68. Fundamentally, such an issue could be due to the non-linear response to the truly observant, slant ozone (not shown here).

$$TCO_{corrected} = TCO_{tempo} \times (5.7e^{-4}\mathrm{SZA} + 0.97)$$

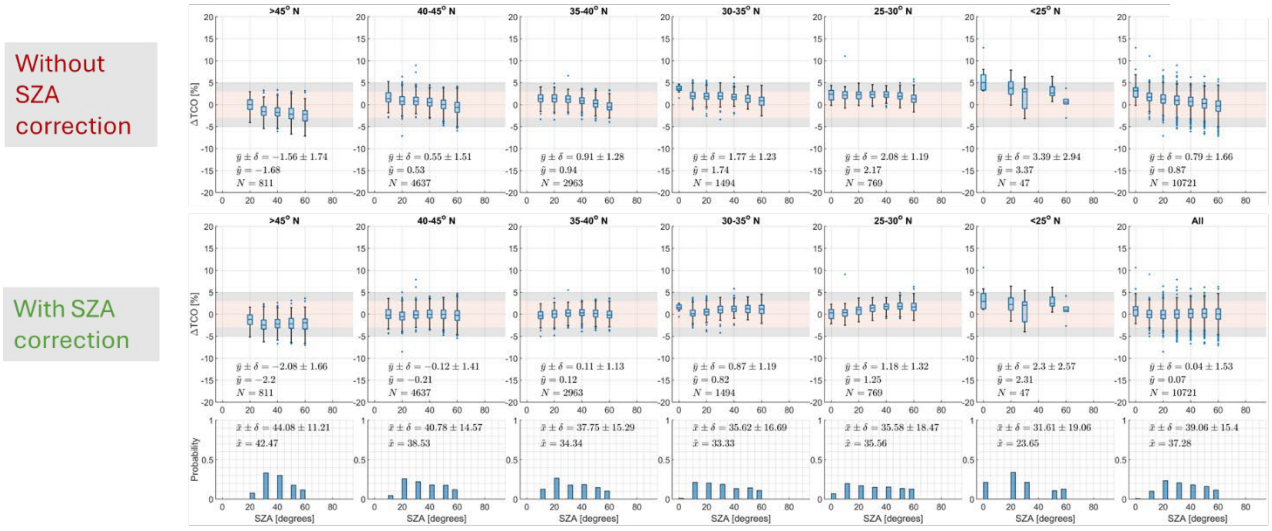


Figure 3-68: Delta Ozone (ΔTCO = (TEMPO-Ground)/Ground) vs. SZA. The first row shows the results without SZA correction; the second row shows the results applied empirical SZA corrections, with parameters fitted using ground-based observations; the third row shows the histogram of SZA values.

Last, TEMPO ozone data shows a VZA dependency when compared with the ground-based networks (not shown here). These results are consistent with findings reported in the TEMPO Total Ozone Level 2 and 3 Data products user guide (which used OMI and OMPS as references). After applying the SZA correction, although the overall percentage difference between TEMPO and ground-based measurements is within -0.23 ± 1.45 % (see Fig. 3-69, first-row last panel), such differences in those six

latitude bins ranged from -1.72% to 2.57% (see Fig. 3-69; from the >45 degrees bins to the <25 degrees bins). Parallel analysis for TROPOMI shows a 0.02 ± 1.34 % overall difference, with very small latitude dependency (see Fig. 3-70). In addition, the results suggest that TEMPO ozone data has slightly lower precision compared to TROPOMI ozone data; the median values of $\sigma\sigma(\Delta TTVVTT)$ are 1.28% and 1.19% for TEMPO and TROPOMI data, respectively.

TEMPO TCO v3 data generally shows reasonable precision and accuracy, with some potential issues that must be addressed in the future. More analysis, such as seasonal (effective temperature) dependency could be examined when more TEMPO observations are available.

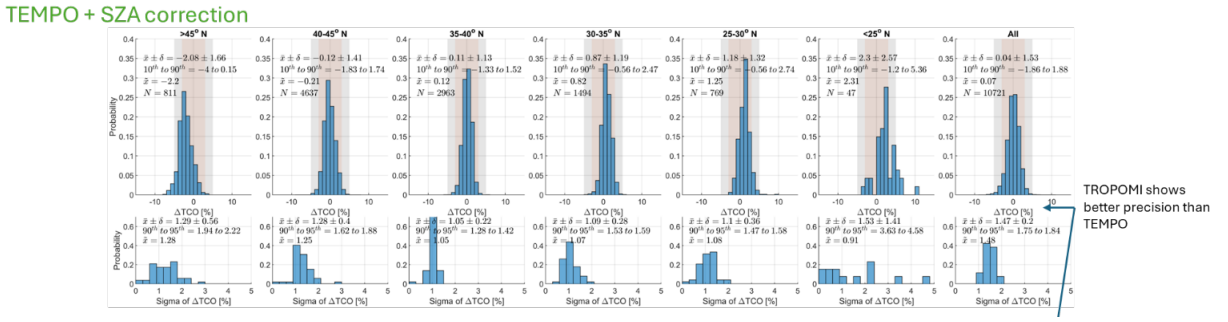


Figure 3-69: Histogram of Delta Ozone and its sigma values for TEMPO (with SZA correction)

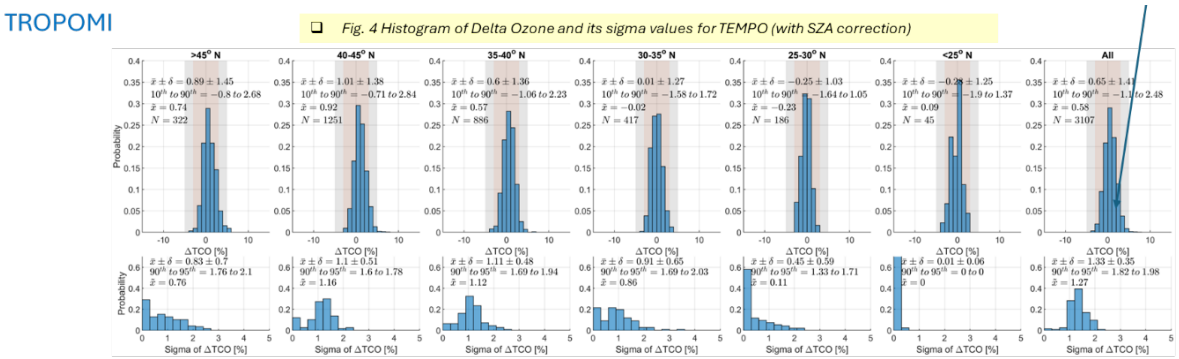


Figure 3-70: Histogram of Delta Ozone and its sigma values for TROPOMI

3.4.3 Validation of TEMPO O_3 V03 with OMI (Ozone Monitoring Instrument), OMPS (Ozone Mapping and Profiler Suite), and Pandora Ground-based spectrometers

Authored by: Jay Herman and Jianping Mao

There are three criteria for successful TEMPO validation of Total Column Ozone (TCO) retrievals. First, is agreement in magnitude within 1% to 3% with other well-calibrated low Earth orbit (LEO) satellite measurements (e.g., OMI Ozone Monitoring Instrument, OMPS Ozone Mapping and Profiler Suite), which are obtained at 13:30±0:20 once every day with an occasional additional measurement 1.5 hours later from the next LEO orbit. Second, is agreement in diurnal variation with accurate ground-based measurements (e.g., the Pandora Spectrometer Network, Brewer Spectrometer Network) and diurnal satellite measurements from a whole Earth synoptic viewing satellite (EPIC Earth Polychromatic Imaging Camera) located near the Earth-Sun Lagrange-1 L_1 gravitational balance point. Third, is a test of internal self-consistency showing that the retrieved ozone values are independent of the solar zenith angle (SZA) and view zenith angle (VZA). The SZA test is easily observable in the data in that the presence of a

parabolic shape roughly centered on noon indicates a calibration error (stray light, radiometric, calibration, or slit-function errors).

There are about 40 operational Pandora spectrometer systems in the field of view (FOV) of TEMPO over the continental US, Canada, and Mexico. This analysis presents a representative subset of comparisons that also include data from polar orbiting satellites OMI (Ozone Monitoring Instrument) onboard the AURA spacecraft and OMPS (Ozone Mapping and Profiler Suite). All of the comparisons are made at locations corresponding to Pandora overpass latitude and longitudes.

Instrument	Data Location
Pandora	https://data.pandonia-global-network.org/
OMI	https://avdc.gsfc.nasa.gov/pub/data/satellite/Aura/OMI/V03/L2OVP/OMTO3/
OMPS	https://acd-ext.gsfc.nasa.gov/anonftp/toms/omps_np/overpass/

At this stage (V03) the goal is to have validation agreement of total column O_3 within about 3%, or a nominal 9 DU based on a typical midlatitude ozone value of 300 DU. There are two types of comparisons needed to validate TEMPO TCO. One is a multi-day time series of Version 3 TEMPO ozone (VO3) with that from well-calibrated Pandora ground-based spectrometer systems and the second is to examine the hourly measurements made by TEMPO in comparison with the minute-by-minute measurements made by Pandora.

Examples of multi-day time series of TCO data are shown from 4 sites Bayonne, New Jersey, Dearborn, Michigan, Richmond, California, and Toronto-Scarborough, Canada using 2-day running averages (Fig. 3-71).

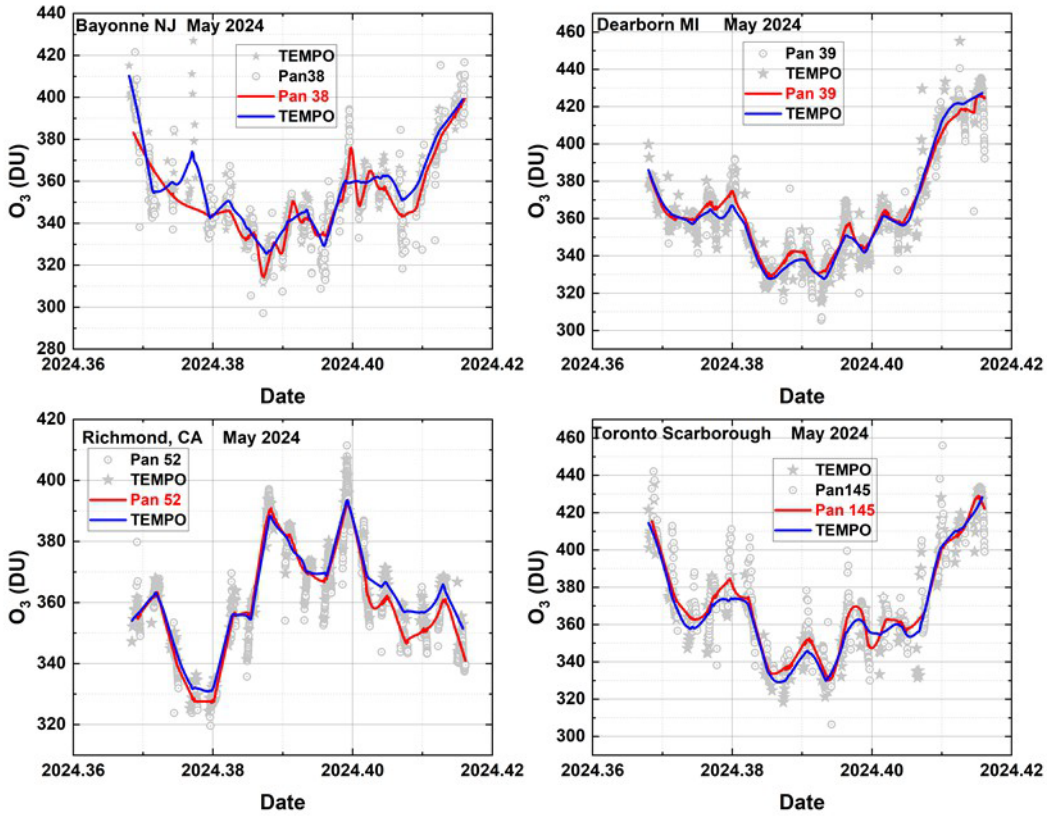


Figure 3-71: Comparison of multi-day time series of TEMPO and Pandora total column ozone at 4 sites. The red and blue lines are 2-day running averages over 18 days.

The agreement is well within the 3% goal and most often within 1 to 2%. This means that long-term total column ozone changes and seasonal variation can be accurately estimated using TEMPO data as the mission progresses.

Examples for measuring diurnal variation of total column ozone TCO using TEMPO are shown in Figs. 3-72, 3-73, 3-74, and 3-75 for overpass of the same 4 Pandora sites. Other well-calibrated Pandora sites have been examined with similar results.

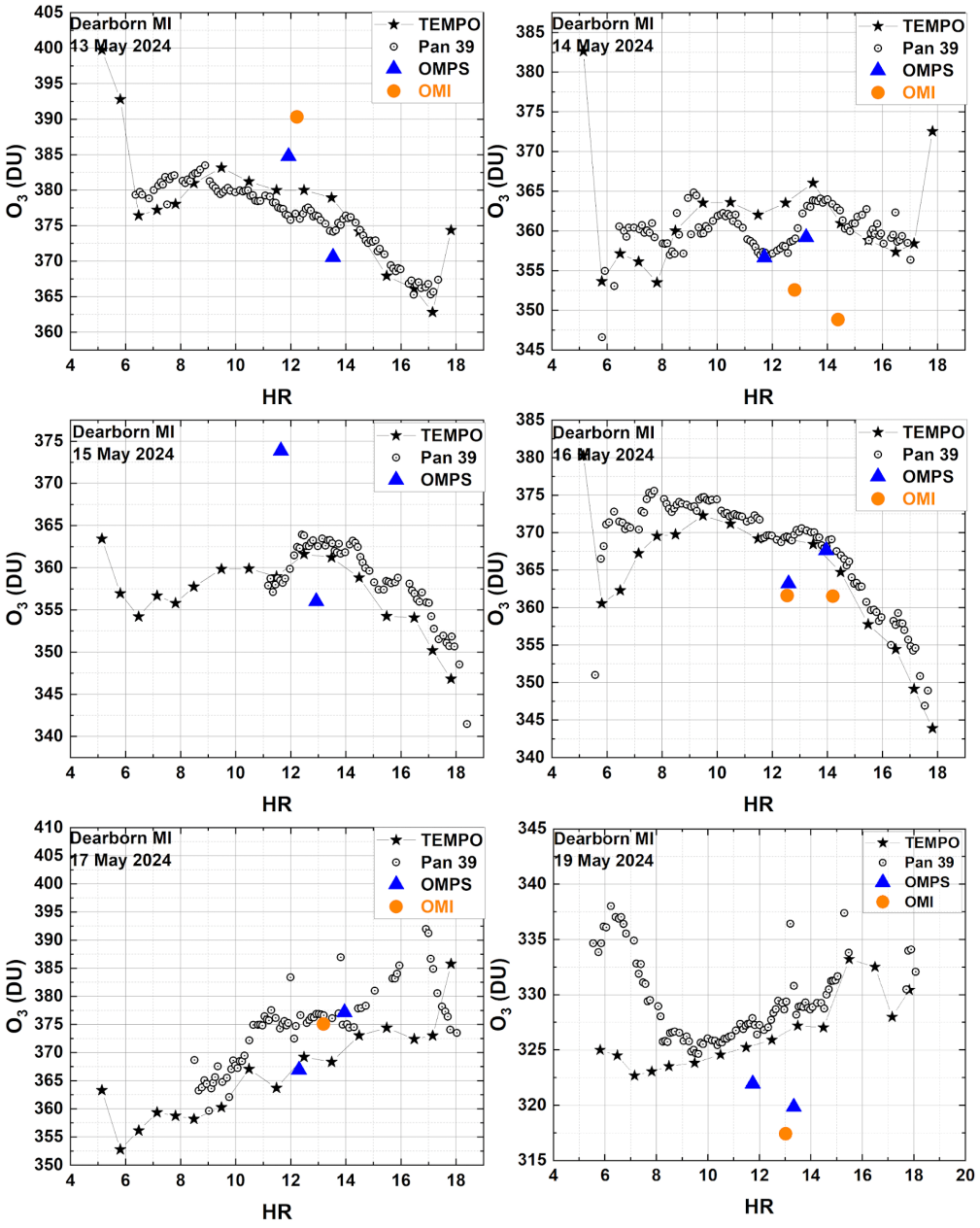


Figure 3-72 Fig. 3.4.3.2 Diurnal TCO variation over Dearborn, Michigan. Stars are TEMPO and circles are Pandora data. Also shown are values from OMPS (Blue Triangles) and OMI (Orange Circles).

The Dearborn Michigan Pandora 39 is a well-maintained and well-calibrated instrument that clearly shows good agreement with TEMPO VO3 TCO on the 6-days shown in Fig. 3-72 and on other days (not shown).

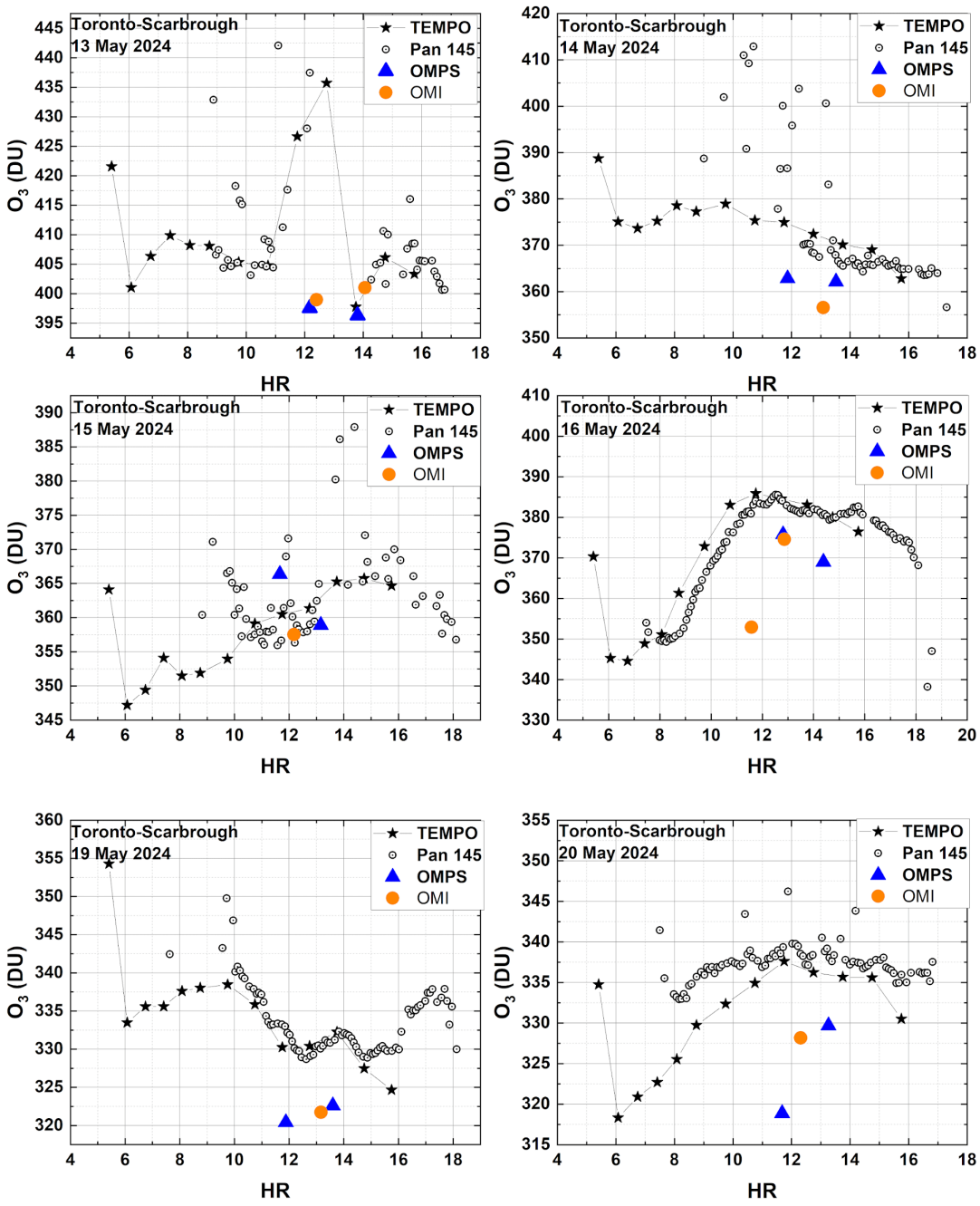


Figure 3-73: Diurnal TCO variation over Toronto-Scarborough. Stars are TEMPO and circles are Pandora data. Also shown are values from OMPS (Blue Triangles) and OMI (Orange Circles).

The TCO comparison for Toronto-Scarborough Canada also shows good agreement of Pandora 145 with TEMPO but here the Pandora data are incomplete on some days.

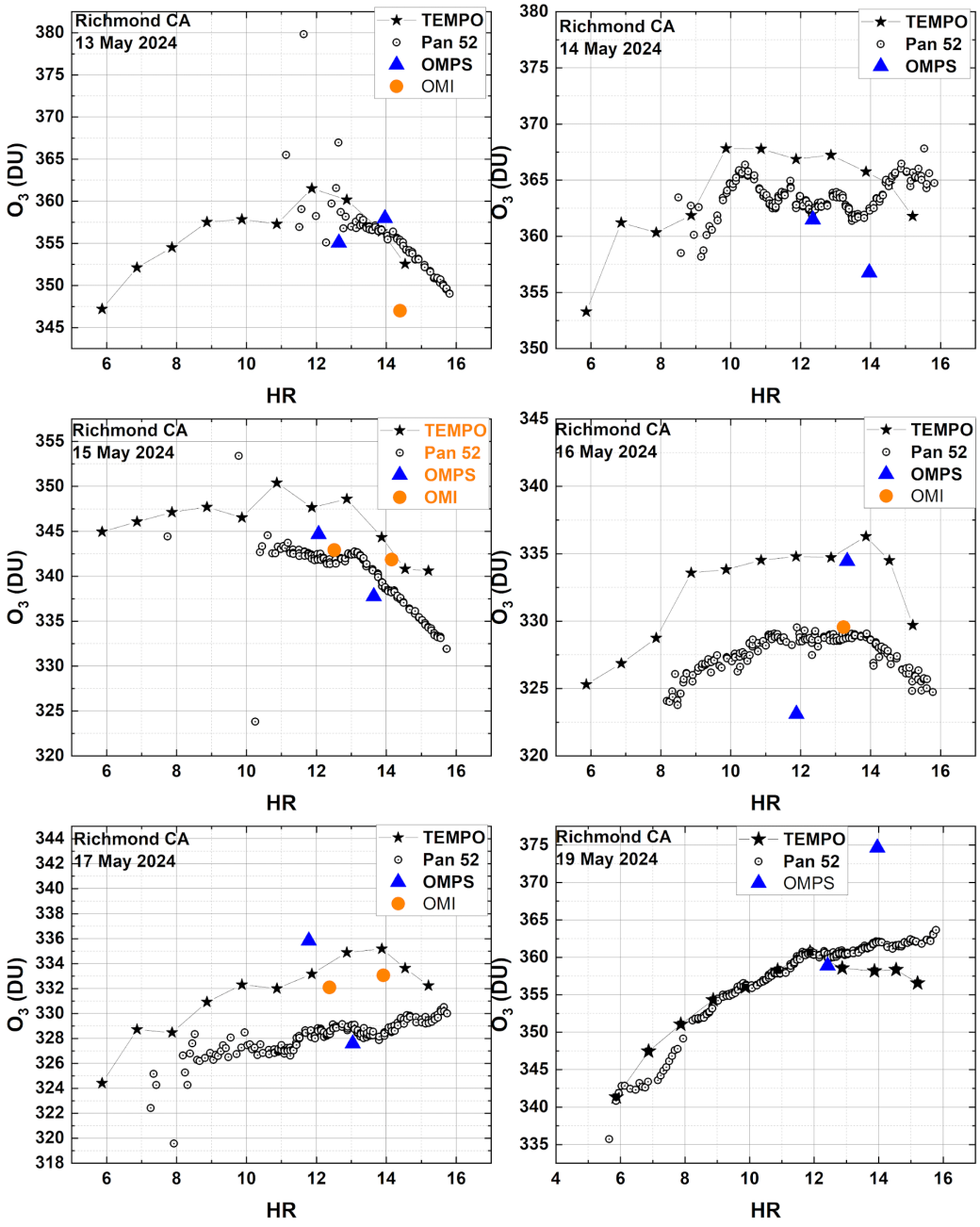


Figure 3-74: Diurnal TCO variation over Richmond California. Stars are TEMPO and circles are Pandora data. Also shown are values from OMPS (Blue Triangles) and OMI (Orange Circles).

The V03 TCO comparison with the Richmond, California Pandora 52 shows some disagreement but well within the goal of 3%.

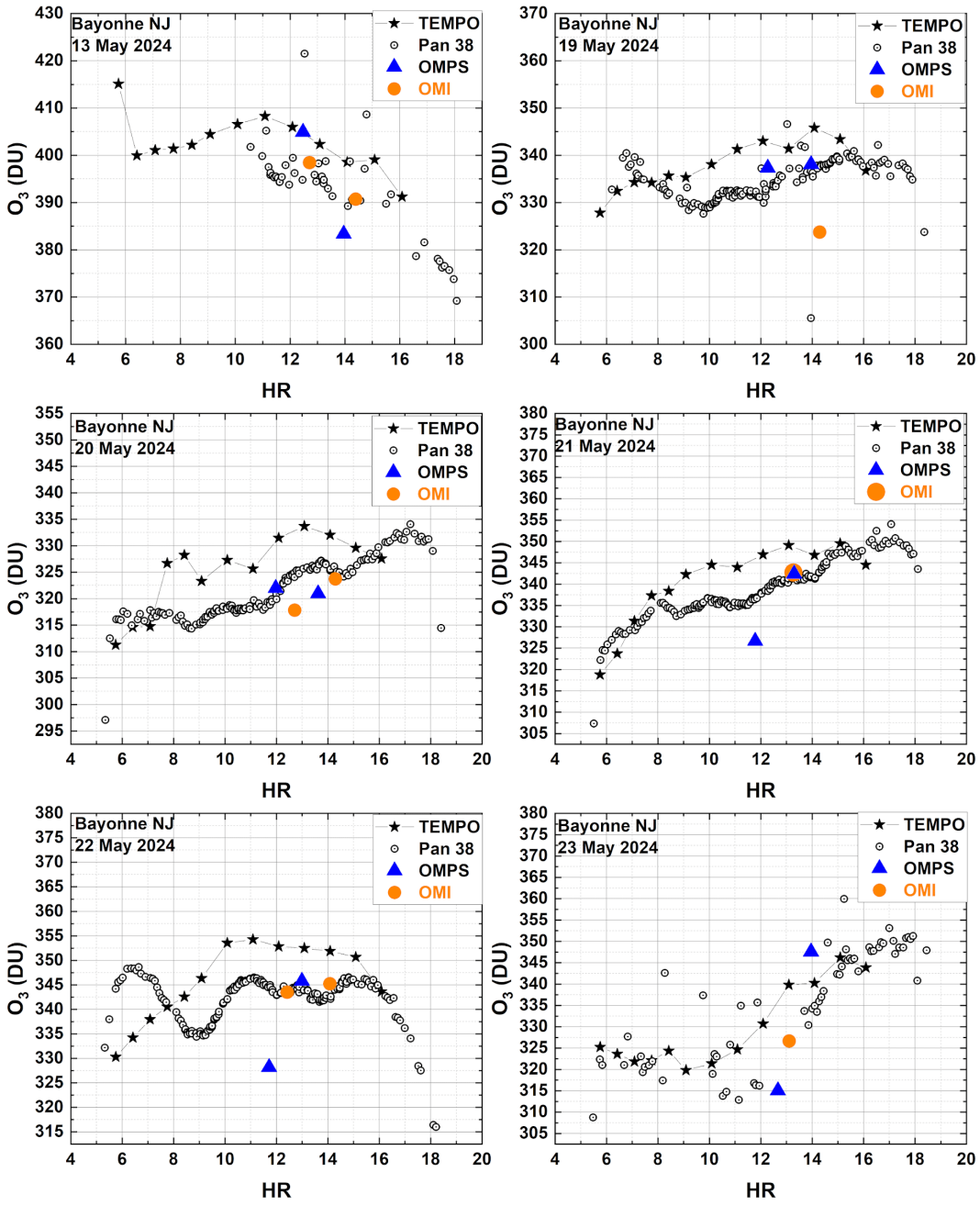


Figure 3-75: Diurnal TCO variation over Bayonne New Jersey. Stars are TEMPO and circles are Pandora data. Also shown are values from OMPS (Blue Triangles) and OMI (Orange Circles).

The agreement between Pandora and TEMPO quite good but sometimes close to 3%. Frequently, the diurnal shape agrees well with Pandora. It should be noted that data obtained early in the morning or late in the afternoon are not expected to agree (high solar zenith angle SZA) because of approximations made for spherical geometry at high SZA for both TEMPO and Pandora.

There also appears to be a latitude dependent bias as shown by comparison with the Mexico City data in Fig. 3.4.3.6. In this example, TCO from DSCOVR EPIC (Earth Polychromatic Imaging Camera) is also shown. Interestingly both disagree with Pandora 142 but in opposite directions

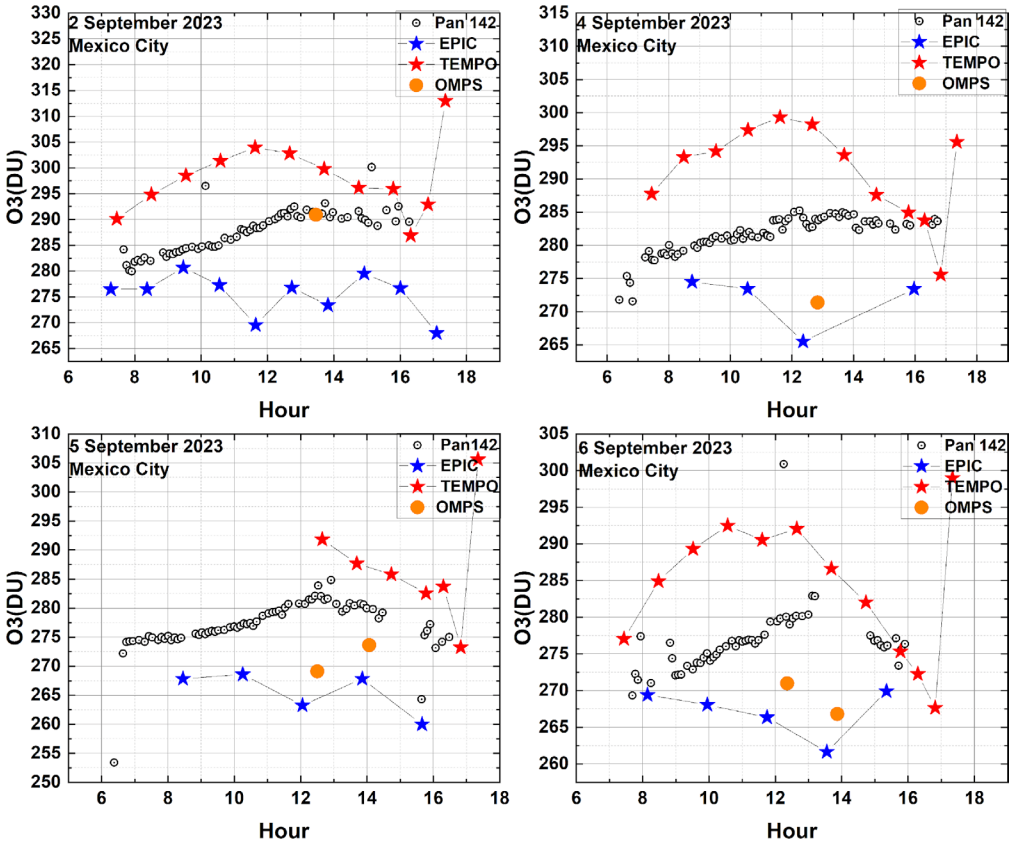


Figure 3-76: Comparison between TEMPO (Red stars), Pandora 142 (circles), DSCOVR EPIC (blue stars), and OMPS (orange circles) over Mexico City, Mexico.

The TEMPO retrievals over the Southern-most city with a Pandora within its field of regard, Mexico City, suggests that TEMPO may have a latitude dependent stray light residual that has not been completely corrected in V03. Stray light errors frequently show up as a parabolic shape centered near solar noon.

3.5 Multi-Species (TropNO2, TotHCHO, TotO3) TEMPO Level 2 Science Product Validation and Evaluation

3.5.1 Background scene assessment of TEMPO , TropNO2, TotHCHO, TotO3

Authored by: Laura Judd (NASA LaRC), Caroline Nowlan, Gonzalo Abad (SAO)

PSPI-03 for all products do not require correlative reference measurements and instead are reflective of the variability of the TEMPO L2 products independently over homogeneous scenes. Here, we present results using TEMPO data from May 30th, 2024, over three homogenous scenes: one bright desert scene over the Desierto del Altar and two water scenes over the Gulf of California and Lake Michigan (0.5º x 0.5º red boxes in Figure 3-77). These scenes did not show significant temporal variance in the L2 products and were 100% cloud free (as confirmed by GOES imagery) except for the final two hours over the Desierto del Altar (16:00-18:00 LT) where a stream of clouds moved over part of the domain.

Figure 3-77 b-d represent statistics on the standard deviation of the L2 products noted, and for NO2 and HCHO, their reported uncertainty gathered over each box during each TEMPO scan on May 30th, 2024. All NO2 and HCHO data have qa_values equal to 0, but for ozone, a qa_value filter was not applied as most pixels had qa_values greater than 0.

For total column ozone, most scans have a standard deviation much less than 1% of the total column. Exceptions to this are the first and last scan over Lake Michigan when the SZA is quite high and when clouds began to impact the Desierto del Altar in the late afternoon. All scans show that this product in these homogeneous scenes achieves better than the required precision of 3% noted in Zoogman et al. (2017).

Tropospheric NO2 VCs are near background values for all scans: 1.0×10^{15} molecules cm^{-2} over Desierto del Altar, 0.5×10^{15} molecules cm^{-2} over the Gulf of California, and 0.7×10^{15} molecules cm^{-2} over Lake Michigan (not shown). The standard deviation of the tropospheric VCs are the same order of magnitude as the column over Lake Michigan (0.9×10^{15} molecules cm^{-2}) and Bay of California (0.6×10^{15} molecules cm^{-2}) but about half the amount over the Desierto del Altar (0.5×10^{15} molecules cm^{-2}). This meets the 1×10^{15} molecules cm^{-2} precision requirement from Zoogman et al. (2017). Qualitatively, the standard deviation of the tropospheric VCs fall within the range of reported uncertainty in the NO2 product. The reported uncertainty of the tropospheric VCs is on average 5% higher than the standard deviation of the tropospheric column over the Desierto del Altar. This value is -2% over the Bay of California and less than 1% over Lake Michigan.

For the HCHO results, the mean of the VCs varies more substantially from region to region with a mean of 8.7×10^{15} molecules cm^{-2} over Desierto del Altar, 4.1×10^{15} molecules cm^{-2} over the Gulf of California, and 2.3×10^{15} molecules cm^{-2} over Lake Michigan (not shown). The standard deviation of the vertical columns over each area are more similar between regions 4.7×10^{15} molecules cm^{-2}, 4.4×10^{15} molecules cm^{-2}, and 5.2×10^{15} molecules cm^{-2}, respectively. This is about half the required precision requirement listed in Zoogman et al. (2017). Similar to NO_2, the standard deviation of the vertical columns fall within the range of reported uncertainty. Over the Desierto del Altar, the reported uncertainty is on average 6% higher than the standard deviation of the vertical column. This value is 13% over the Gulf of California and 9% over Lake Michigan. This makes the uncertainty estimate larger than the observed variation of the columns in these cases.

While the precision requirement is met over homogeneous scenes, more work is needed to assess the pixel-to-pixel variance over more heterogeneous scenes with respect to surface characteristics, clouds, and pollution gradients.

The grey bars in each graph indicate the maximum cloud effective fraction over the area. Hours that show an increase in effective cloud fraction also show an increase in the uncertainty and the standard deviation of the columns for both products. Note that all areas have cloud effective fractions maximizing above 0 even in cloud free scenes with even higher fractions over the bright desert case. This is not what is occurring in reality. The impact of the cloud retrieval on the trace gas retrievals should be assessed and is an area of improvement needed for future retrieval versions.

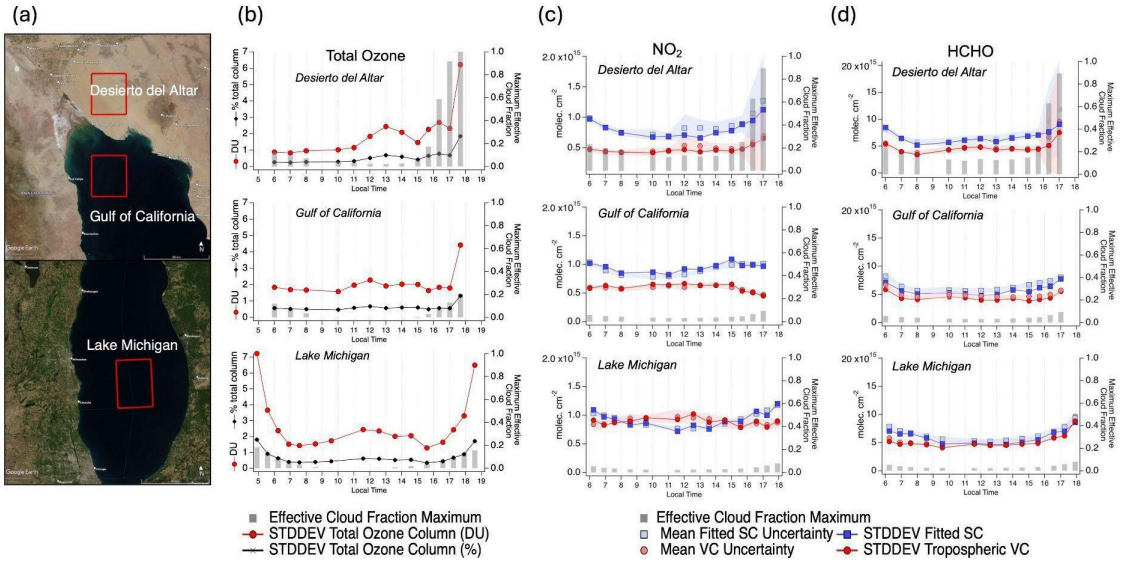


Figure 3-77: (a)VIIRS true color imagery from Worldview on May 20th, 2024 with the Desierto del Altar, Gulf of California, and Lake Michigan regions outlined in red. (b) Time series of the standard deviation of the TEMPO Total ozone column (DU in red, % in black) within each region. (c-d) Time series of the standard deviation of the NO2 tropospheric vertical column and HCHO total column (dark red) and fitted slant columns (dark blue) for each hour within each region. The lighter red and blue markers and shading indicate the mean+/- standard deviation of uncertainty reported for each parameter in each region. (b-d) grey bars indicate the maximum effective cloud fraction reported in each region for each hour. All times are in local time for each region.

4. References, Acknowledgements, and Data Availability

References:

Baidar, S., Oetjen, H., Coburn, S., Dix, B., Ortega, I., Sinreich, R., and Volkamer, R.: The CU Airborne MAX-DOAS instrument: vertical profiling of aerosol extinction and trace gases, Atmospheric Measurement Techniques, 6, 719–739, https://doi.org/10.5194/amt-6-719-2013, 2013.

Bovensmann, H., Burrows, J.P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V.V., Chance, K.V. and Goede, A.P.H., 1999. SCIAMACHY: Mission objectives and measurement modes. Journal of the atmospheric sciences, 56(2), pp.127-150.

Bucsela, E. J., Perring, A. E., Cohen, R. C., Boersma, K. F., Celarier, E. A., Gleason, J. F., Wenig, M. O., Bertram, T. H., Wooldridge, P. J., Dirksen, R., and Veefkind, J. P.: Comparison of tropospheric NO 2 from

in situ aircraft measurements with near-real-time and standard product data from OMI, Journal of Geophysical Research, 113, https://doi.org/10.1029/2007JD008838, 2008.

Burrows, J.P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., DeBeek, R., Hoogen, R., Bramstedt, K., Eichmann, K.U. and Eisinger, M., 1999. The global ozone monitoring experiment (GOME): Mission concept and first scientific results. Journal of the Atmospheric Sciences, 56(2), pp.151-175.

Cazorla, M., G. M. Wolfe, S. A. Bailey, A. K. Swanson, H. L. Arkinson, and T. F. Hanisco: A new airborne laser-induced fluorescence instrument for in situ detection of formaldehyde throughout the troposphere and lower stratosphere, Atmospheric Measurement Techniques, 8, 541-552, www.atmos-meas-tech.net/8/541/2015/, 2015.

Cheng, Siyang, Cheng, Xinghong, Ma, Jianzhong, Xu, Xiangde, Zhang, Wenqian, Lv, Jinguang, Bai, Gang, Chen, Bing, Ma, Siying, Ziegler, Steffen, Donner, Sebastian, Wagner, Thomas. 2023. Mobile MAX-DOAS observations of tropospheric NO2 and HCHO during summer over the Three Rivers' Source region in China. Atmospheric Chemistry and Physics. Vol. 23, No. 6, pp. 3655-3677. DOI: 10.5194/acp-23-3655-2023 ISSN: 1680-7324

Copernicus Sentinel data processed by ESA, German Aerospace Center (DLR) (2020), Sentinel-5P TROPOMI Tropospheric Formaldehyde HCHO 1-Orbit L2 5.5km x 3.5km, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 2024-11-23, 10.5270/S5P-vg1i7t0

Copernicus Sentinel data processed by ESA, Koninklijk Nederlands Meteorologisch Instituut (KNMI) (2021), Sentinel-5P TROPOMI Tropospheric NO2 1-Orbit L2 5.5km x 3.5km, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 2024-11-23, 10.5270/S5P-9bnp8q8

Flynn, L., Long, C., Wu, X., Evans, R., Beck, C.T., Petropavlovskikh, I., McConville, G., Yu, W., Zhang, Z., Niu, J. and Beach, E., 2014. Performance of the ozone mapping and profiler suite (OMPS) products. Journal of Geophysical Research: Atmospheres, 119(10), pp.6181-6195.

Judd, L. M., Al-Saadi, J. A., Szykman, J. J., Valin, L. C., Janz, S. J., Kowalewski, M. G., Eskes, H. J., Veefkind, J. P., Cede, A., Mueller, M., Gebetsberger, M., Swap, R., Pierce, R. B., Nowlan, C. R., Abad, G. G., Nehrir, A., and Williams, D.: Evaluating Sentinel-5P TROPOMI tropospheric NO2 column densities with airborne and Pandora spectrometers near New York City and Long Island Sound, Atmospheric Measurement Techniques, 13, 6113–6140, https://doi.org/10.5194/amt-13-6113-2020, 2020.

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., Lucchesi, R. A., Lundgren, E. W., Nicely, J. M., Nielsen, E., Ott, L. E., Saunders, E., Strode, S. A., Wales, P. A., Jacob, D. J., and Pawson, S.: Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0, J. Adv. Model. Earth Syst., 13, https://doi.org/10.1029/2020MS002413, 2021.

Knowland, K. E., Keller, C. A., Wales, P. A., Wargan, K., Coy, L., Johnson, M. S., Liu, J., Lucchesi, R. A., Eastham, S. D., Fleming, E., Liang, Q., Leblanc, T., Livesey, N. J., Walker, K. A., Ott, L. E., and Pawson, S.: NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0: Stratospheric Composition, J Adv Model Earth Syst, 14, e2021MS002852, https://doi.org/10.1029/2021MS002852, 2022.

Lambert, J.-C., A. Keppens, S. Compernolle, K.-U. Eichmann, M. de Graaf, D. Hubert, B. Langerock, A. Ludewig, M.K. Sha, T. Verhoelst, T. Wagner, C. Ahn, A. Argyrouli, D. Balis, K.L. Chan, M. Coldewey-Egbers, I. De Smedt, H. Eskes, A.M. Fjæraa, K. Garane, J.F. Gleason, F. Goutail, J. Granville, P. Hedelt, K.-P. Heue, G. Jaross, Q. Kleipool, ML. Koukouli, R. Lutz, M.C Martinez Velarte, K. Michailidis, A. Pseftogkas, S. Nanda, S. Niemeijer, A. Pazmiño, G. Pinardi, A. Richter, N. Rozemeijer, M. Sneep, D. Stein Zweers, N. Theys, G. Tilstra, O. Torres, P. Valks, J. van Geffen, C. Vigouroux, P. Wang, and M. Weber, Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products #24: April 2018 – August 2024. S5P MPC Routine Operations Consolidated Validation Report series, Issue #24, Version 24.00.00, 212 pp., 16 September 2024

Levelt, P.F., Joiner, J., Tamminen, J., Veefkind, J.P., Bhartia, P.K., Stein Zweers, D.C., Duncan, B.N., Streets, D.G., Eskes, H., van der A, R. and McLinden, C., 2018. The Ozone Monitoring Instrument: overview of 14 years in space. Atmospheric Chemistry and Physics, 18(8), pp.5699-5745.

Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., Huckle, R., Lacan, A., Grzegorski, M., Holdak, A. and Kokhanovsky, A., 2016. The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing–an overview. Atmospheric Measurement Techniques, 9(3), pp.1279-1301.

National Aeronautics and Space Administration (2023), Tropospheric Emissions: Monitoring of Pollution (TEMPO) Project, Level 2 Science Data Product Validation Plan SAO-DRD-11, 25 APRIL 2023, https://tempo.si.edu/documents/SAO-DRD-11_TEMPO%20Science%20Validation_Plan_Baseline.pdf

Nawaz, M. O., Johnson, J., Yarwood, G., de Foy, B., Judd, L., and Goldberg, D. L.: An intercomparison of satellite, airborne, and ground-level observations with WRF–CAMx simulations of NO2 columns over Houston, Texas, during the September 2021 TRACER-AQ campaign, Atmospheric Chemistry and Physics, 24, 6719–6741, https://doi.org/10.5194/acp-24-6719-2024, 2024.

Oetjen, H., S. Baidar, N. A. Krotkov, L. N. Lamsal, M. Lechner, and R. Volkamer (2013), Airborne MAX-DOAS measurements over California: Testing the NASA OMI tropospheric NO 2 product, J. Geophys. Res. Atmos., 118, 7400–7413, doi:10.1002/jgrd.50550

Rawat, P., Crawford, J. H., Travis, K. R., Judd, L. M., Demetillo, M. A. G., Valin, L. C., Szykman, J. J., Whitehill, A., Baumann, E., and Hanisco, T. F.: Maximizing the Use of Pandora Data for Scientific Applications, Atmos. Meas. Tech., 18, 2899–2917, https://doi.org/10.5194/amt-18-2899-2025, 2025.

Rollins, A. W., Rickly, P. S., Gao, R.-S., Ryerson, T. B., Brown, S. S., Peischl, J., and Bourgeois, I.: Single-photon laser-induced fluorescence detection of nitric oxide at sub-parts-per-trillion mixing ratios, Atmospheric Measurement Techniques, 13, 2425–2439, https://doi.org/10.5194/amt-13-2425-2020, 2020.

Vigouroux, C., Bauer Aquino, C. A., Bauwens, M., Becker, C., Blumenstock, T., De Mazière, M., García, O., Grutter, M., Guarin, C., Hannigan, J., Hase, F., Jones, N., Kivi, R., Koshelev, D., Langerock, B., Lutsch, E., Makarova, M., Metzger, J.-M., Müller, J.-F., Notholt, J., Ortega, I., Palm, M., Paton-Walsh, C., Poberovskii, A., Rettinger, M., Robinson, J., Smale, D., Stavrakou, T., Stremme, W., Strong, K., Sussmann, R., Té, Y., and Toon, G.: NDACC harmonized formaldehyde time series from 21 FTIR stations covering a wide range of column abundances, Atmos. Meas. Tech., 11, 5049–5073, https://doi.org/10.5194/amt-11-5049-2018, 2018.

Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., Cheng, Z., De Mazière, M., De Smedt, I., Grutter, M., Hannigan, J. W., Jones, N., Kivi, R., Loyola, D., Lutsch, E., Mahieu, E., Makarova, M., Metzger, J.-M., Morino, I., Murata, I., Nagahama, T., Notholt, J., Ortega, I., Palm, M., Pinardi, G., Röhling, A., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., van Roozendael, M., Wang, P., and Winkler, H.: TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations, Atmos. Meas. Tech., 13, 3751–3767, https://doi.org/10.5194/amt-13-3751-2020, 2020.

Zhu, L., et al., Validation of satellite formaldehyde (HCHO) retrievals using observations from 12 aircraft campaigns, Atmos. Chem. Phys., 20, 12329–12345, https://doi.org/10.5194/acp-20-12329-2020, 2020.

Zoogman, P., Liu, X., Suleiman, R. M., Pennington, W. F., Flittner, D. E., Al-Saadi, J. A., Hilton, B. B., Nicks, D. K., Newchurch, M. J., Carr, J. L., Janz, S. J., Andraschko, M. R., Arola, A., Baker, B. D., Canova, B. P., Chan Miller, C., Cohen, R. C., Davis, J. E., Dussault, M. E., Edwards, D. P., Fishman, J., Ghulam, A., González Abad, G., Grutter, M., Herman, J. R., Houck, J., Jacob, D. J., Joiner, J., Kerridge, B. J., Kim, J., Krotkov, N. A., Lamsal, L., Li, C., Lindfors, A., Martin, R. V., McElroy, C. T., McLinden, C., Natraj, V., Neil, D. O., Nowlan, C. R., O'Sullivan, E. J., Palmer, P. I., Pierce, R. B., Pippin, M. R., Saiz-Lopez, A., Spurr, R. J. D., Szykman, J. J., Torres, O., Veefkind, J. P., Veihelmann, B., Wang, H., Wang, J., and Chance, K.: Tropospheric emissions: Monitoring of pollution (TEMPO), Journal of Quantitative Spectroscopy and Radiative Transfer, 186, 17–39, https://doi.org/10.1016/j.jqsrt.2016.05.008, 2017.

Acknowledgements:

We thank the PGN Principal Investigators, support staff and funding for establishing and maintaining the PGN sites used in this investigation. The PGN is a bilateral project supported with funding from NASA and ESA with additional support provided by the U.S. EPA to maintain network operations and create an expanded network under the TEMPO FOR.

Data Availability Statement:

The data supporting the findings of this validation report are derived from public domain resources including ground-based networks, atmospheric composition field campaigns, and satellite observations. The specific datasets are categorized as follows:

1. TEMPO Satellite Observations

The **TEMPO Level 1 (radiances)** and **Level 2 (trace gas columns)** version 3 data products (NO_2, O_3, and HCHO) are available through the NASA Earthdata Search portal under the TEMPO project collection.

2. Validation Reference Data

To assess retrieval accuracy, this report utilizes independent measurements from the following sources:

- **Ground-Based Network Observations:**
    - **Pandonia Global Network (PGN):** High-frequency direct-sun measurements from Pandora spectrometers serve as the primary validation standard. Available at pandonia-global-network.org.
    - **FTIR Network (NDACC):** High-resolution solar absorption spectra for trace gas profiles are sourced from the Network for the Detection of Atmospheric Composition Change (NDACC). North American site data are available via the NDACC Data Host Facility.

  - **Brewer and Dobson Spectrometers:** Total column ozone (O_3) validation relies on the global network of ultraviolet spectrophotometers. Data are available via the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) at woudc.org.

- **Intensive Field Campaign Datasets (Summer 2023):** This report leverages synergistic observations from the **AGES+** multi-agency initiative (AEROMMA, CUPIDS, GOTHAAM, and STAQS), providing high-resolution aircraft and ground measurements:

  - **STAQS:** Airborne remote sensing (GCAS) and ground-based lidar (TOLNet) data are archived at the NASA STAQS Data Archive.
  - **AEROMMA:** In situ trace gas and aerosol measurements from the NASA DC-8 are available at the NOAA CSL AEROMMA Data Archive.
  - **CUPIDS:** Boundary layer dynamics and transport data from the NOAA Twin Otter and ground sites are hosted at the NOAA CSL CUPIDS Archive.

- **Satellite & Model Datasets:**
  - **TROPOMI:** Accessed through the Copernicus Data Space Ecosystem.
  - **OMI:** Accessed through the NASA Earthdata Search portal.
  - **WRF-Chem Model:** University of Wisconsin-Madison 4km retrospective chemistry and aerosol predictions are available via the NOAA CSL AEROMMA Data Archive.