



Sentinel-5P Data Pipeline: Acquisition, Processing, and Analysis for Global Atmospheric Monitoring

Sayali Shripad Mahulkar*

Research Student, Department of Statistics, Institute of Science, Civil Lines, Nagpur, Maharashtra, India 440001.

Open Access Research Journal of Science and Technology, 2026, 16(01), 031-038

Publication history: Received on 20 December 2025; revised on 28 January 2026; accepted on 30 January 2026

Article DOI: <https://doi.org/10.53022/oarjst.2026.16.1.0007>

Abstract

Global atmospheric monitoring requires integrated workflows that transform raw satellite observations into actionable environmental intelligence. The Sentinel-5 Precursor (S5P) satellite mission, equipped with the TROPospheric Monitoring Instrument (TROPOMI), provides unprecedented global coverage of atmospheric trace gases at high spatial and temporal resolution. However, the complexity of satellite data acquisition, preprocessing, quality control, and analytical processing presents substantial barriers to effective utilization, particularly for researchers in resource-constrained environments. This paper presents a comprehensive end-to-end data pipeline for Sentinel-5P atmospheric monitoring, encompassing automated data acquisition through Application Programming Interfaces (APIs), systematic preprocessing workflows, rigorous quality assurance procedures, integration with auxiliary meteorological datasets, advanced analytical techniques, and validation methodologies. The pipeline architecture emphasizes modularity, reproducibility, and scalability, enabling researchers to process multi-year datasets across multiple pollutant species including nitrogen dioxide (NO₂), sulfur dioxide (SO₂), carbon monoxide (CO), and ozone (O₃). Specific focus is given to automation strategies that minimize manual intervention, error handling mechanisms that ensure data integrity, and computational optimization techniques that enable processing on standard research computing infrastructure. The framework leverages open-source tools including Python scientific libraries, cloud-based data access systems, and modern data engineering practices to democratize access to satellite-based atmospheric monitoring capabilities. By providing detailed technical specifications, implementation guidance, and best practices, this pipeline enables the global research community to harness Sentinel-5P data for air quality assessment, emission monitoring, atmospheric chemistry research, and policy-relevant environmental analysis[1][2][3].

Keywords: Sentinel-5P; TROPOMI; Data pipeline; Atmospheric monitoring; Automated processing; Quality assurance; Satellite data analysis; Air pollution

1. Introduction

1.1. The Challenge of Global Atmospheric Monitoring

Atmospheric pollution represents a critical global environmental challenge with profound implications for human health, ecosystem integrity, and climate dynamics. The World Health Organization attributes approximately 7 million premature deaths annually to ambient air pollution, with nitrogen dioxide (NO₂), sulfur dioxide (SO₂), particulate matter, and ground-level ozone (O₃) serving as primary contributors to this burden[4]. Despite this enormous public health impact, comprehensive atmospheric monitoring remains severely limited in spatial coverage, particularly in low- and middle-income regions where pollution burdens are often highest but monitoring infrastructure is most constrained[5].

Traditional ground-based monitoring networks, while providing high-accuracy point measurements, suffer from fundamental limitations in spatial representativeness. Monitoring stations are expensive to establish and maintain,

* Corresponding author: Sayali Shripad Mahulkar

requiring continuous power supply, technical expertise, regular calibration, and quality control procedures. Consequently, ground-based networks exhibit substantial geographic bias, with dense coverage in wealthy urban areas of developed nations contrasting sharply with sparse or nonexistent coverage across vast regions of Africa, South Asia, and Latin America[6]. This monitoring gap creates a critical knowledge deficit precisely where pollution exposures may be most severe and policy interventions most urgently needed.

1.2. Satellite Remote Sensing as a Transformative Solution

Satellite remote sensing offers a paradigm shift in atmospheric monitoring by providing spatially continuous observations with global coverage and consistent measurement protocols. The Sentinel-5 Precursor (S5P) satellite, launched on October 13, 2017, as part of the European Union's Copernicus Programme, carries the TROPospheric Monitoring Instrument (TROPOMI), representing the state-of-the-art in atmospheric composition monitoring from space[7][8]. TROPOMI provides daily global observations of key atmospheric trace gases including NO₂, SO₂, CO, O₃, methane (CH₄), and formaldehyde (HCHO), along with aerosol properties and cloud characteristics.

The transformative potential of TROPOMI derives from its unprecedented combination of spatial resolution (5.5 km × 3.5 km for most products, with enhanced resolution capabilities approaching 1 km² for specialized applications), temporal resolution (daily global coverage with afternoon overpass), and spectral coverage (ultraviolet through shortwave-infrared wavelengths enabling simultaneous multi-species retrieval)[2]. These characteristics enable identification of individual emission sources, tracking of pollution transport patterns, and quantification of atmospheric composition at scales directly relevant to urban air quality management and regulatory decision-making.

1.3. The Data Pipeline Challenge

While TROPOMI's technical capabilities are exceptional, realizing the full potential of satellite atmospheric monitoring requires overcoming substantial data engineering and analytical challenges. The complete workflow from raw satellite observations to policy-relevant environmental intelligence encompasses multiple complex steps[9][10]:

- **Data Discovery and Acquisition:** Identifying relevant satellite products, constructing appropriate queries, authenticating with data distribution systems, and managing bulk downloads of multi-terabyte datasets
- **Data Extraction and Preprocessing:** Parsing hierarchical NetCDF file structures, extracting geophysical variables, applying coordinate transformations, and handling missing data
- **Quality Assurance:** Implementing multi-stage quality filtering based on retrieval confidence, cloud contamination, surface properties, and algorithmic performance indicators
- **Data Integration:** Combining satellite observations with auxiliary datasets including meteorological reanalysis products, topographic information, and land cover characteristics
- **Analytical Processing:** Computing derived quantities, performing spatial and temporal aggregations, executing statistical analyses, and generating visualization products
- **Validation and Uncertainty Quantification:** Comparing satellite estimates with ground-truth measurements, quantifying systematic biases, and characterizing uncertainty propagation

Each of these steps requires specialized technical knowledge spanning remote sensing physics, atmospheric science, data engineering, statistical analysis, and scientific programming. The learning curve presents a substantial barrier to entry, particularly for early-career researchers, scientists in resource-constrained institutions, and policy-oriented analysts without extensive remote sensing backgrounds.

1.4. Objective and Scope

This paper presents a comprehensive data pipeline for Sentinel-5P atmospheric monitoring that addresses these challenges through systematic automation, robust error handling, and modular architecture. The pipeline is designed to be[11][12]:

- **Comprehensive:** Covering the complete workflow from initial data discovery through final validated analytical products
- **Automated:** Minimizing manual intervention through scripted workflows, batch processing, and intelligent error recovery
- **Reproducible:** Employing version-controlled code, documented parameters, and standardized file formats to ensure research reproducibility

- **Scalable:** Supporting processing of multi-year, multi-region, multi-pollutant datasets on standard research computing infrastructure
- **Modular:** Enabling component reuse, customization, and integration with existing analytical workflows
- **Open and Accessible:** Built entirely on open-source tools and open-access data, eliminating cost barriers to adoption

By providing detailed technical specifications, reference implementations, and best practices, this work aims to democratize access to satellite-based atmospheric monitoring capabilities, enabling the global research community to leverage TROPOMI data for air quality assessment, emission source characterization, atmospheric chemistry research, and evidence-based policy development.

2. Pipeline Architecture and Design Principles

2.1. Modular Pipeline Design

The Sentinel-5P data pipeline employs a modular architecture consisting of six primary components, each implementing a discrete processing stage with well-defined inputs, outputs, and error handling mechanisms. This modular approach enables independent testing, parallel execution, and flexible customization for specific research applications[13].

The six core modules are:

- **Module 1: Data Discovery and Query Construction:** Identifies available satellite products matching temporal, spatial, and product-type criteria through programmatic interaction with the Copernicus Data Space Ecosystem (CDSE) catalogue API
- **Module 2: Authenticated Data Acquisition:** Manages OAuth 2.0 authentication, implements parallel download workflows, performs integrity verification, and handles network resilience
- **Module 3: NetCDF Extraction and Quality Filtering:** Parses hierarchical NetCDF structures, extracts geophysical variables, applies multi-stage quality assurance filters, and exports to analysis-ready formats
- **Module 4: Auxiliary Data Integration:** Retrieves and integrates meteorological reanalysis data (ERA5 Planetary Boundary Layer Height, surface meteorology), land cover information, and topographic data
- **Module 5: Analytical Processing and Derived Products:** Computes spatial aggregations, temporal statistics, trend analyses, and derived atmospheric quantities
- **Module 6: Validation and Quality Assessment:** Performs statistical validation, uncertainty quantification, comparison with ground-truth measurements, and generates quality assessment reports

2.2. Design Principles

The pipeline architecture adheres to established software engineering best practices adapted for scientific computing workflows[14][15]:

- **Separation of Concerns:** Each module implements a single, well-defined responsibility with minimal coupling to other components. Data exchange between modules occurs through standardized file formats (CSV, GeoJSON, NetCDF) rather than in-memory objects, enabling check pointing, restart capability, and debugging.
- **Error Handling and Resilience:** All modules implement comprehensive exception handling with three-tier response strategies: (1) automatic retry with exponential backoff for transient failures (network timeouts, rate limiting), (2) graceful degradation for partial failures (processing subset of available data when some files are corrupted), and (3) informative failure reporting for unrecoverable errors with detailed logging for debugging.
- **Configuration Management:** All processing parameters (quality thresholds, spatial domains, temporal ranges, file paths) are specified in external configuration files (YAML or JSON format) rather than hard-coded, enabling parameter sweeps, sensitivity analyses, and reproducible research workflows.
- **Logging and Provenance:** Each module generates structured logs recording processing timestamps, input file identifiers, parameter values, quality metrics, and success/failure status. These logs enable complete provenance tracking from raw satellite observations to final analytical products.
- **Computational Efficiency:** Processing workflows exploit parallelization opportunities including concurrent file downloads, parallel NetCDF parsing, and vectored numerical operations using NumPy and Pandas libraries. Memory management strategies prevent excessive RAM consumption when processing large multi-year datasets.

2.3. Data Discovery And Query Construction

The Copernicus Data Space Ecosystem (CDSE) represents a modern cloud-native data distribution platform offering a Catalogue API for searching available products, a Download API for authenticated file retrieval, and a Visualization API for on-demand data rendering[3][17]. Access requires free registration at <https://dataspace.copernicus.eu/> and OAuth 2.0 authentication[18].

The Catalogue API employs OData (Open Data Protocol) query syntax enabling complex filtering by collection name, product type, temporal range, and geographic region[19]. Geographic filtering uses Well-Known Text POLYGON format with longitude-latitude coordinates. A single city-scale region over one year typically yields 300-400 satellite overpasses per pollutant. The platform returns paginated results with up to 1,000 products per page, requiring systematic pagination handling for large queries.

Recommended practices for efficient acquisition include: (1) temporal chunking breaking annual queries into quarterly segments, (2) parallel downloads implementing 4-8 concurrent threads to maximize bandwidth while respecting rate limits, (3) SHA-256 checksum verification using CDSE-provided values to confirm download integrity, and (4) metadata logging recording timestamp, file size, checksum, and success status for complete audit trails[3].

2.4. Pollutant-Specific Data Products

Table 1 The four primary air quality-relevant L2 data products

Product Code	Pollutant	Column Variable	Units	Scientific Importance
OFFL_L2_NO2__	Nitrogen Dioxide	nitrogen_dioxide_tropospheric_column	mol/m ²	Urban emissions, photochemical smog precursor
OFFL_L2_SO2__	Sulfur Dioxide	sulfur_dioxide_total_vertical_column	mol/m ²	Volcanic emissions, industrial pollution indicator
OFFL_L2_CO__	Carbon Monoxide	carbonmonoxide_total_column	mol/m ²	Combustion tracer, health hazard indicator
OFFL_L2_O3__	Ozone	ozone_total_vvertical_column	mol/m ²	Stratospheric/tropospheric ozone layer indicator

Each product is delivered as a NetCDF-4 hierarchical file containing measurements from a single satellite overpass (approximately 10-minute observation window). The hierarchical structure separates measurement data, quality metrics, and supporting information into organized groups, enabling efficient data access and processing.

3. Data Acquisition and Processing

3.1. NetCDF Processing

Sentinel-5P Level-2 products are distributed as NetCDF-4 hierarchical files containing a complete single satellite overpass observation (~10 minutes). The PRODUCT group contains geophysical retrievals including latitude, longitude, column density measurements, and quality assurance flags (qa_value, 0-1 scale)[21].

Quality Assurance: Multi-stage filtering progressively removes unreliable retrievals. Stage 1 applies qa_value filtering with recommended threshold ≥ 0.75 for high-confidence applications, retaining 40-60% of pixels. Stage 2 removes non-physical values (negative columns, zero values from failed algorithms). Stage 3 identifies extreme outliers ($>5\sigma$ from regional climatology) indicating potential corruption[22][23].

Spatial Reference: TROPOMI measurements use WGS84 geographic coordinates. Consistent maintenance of the same coordinate reference system throughout processing is critical to avoid accumulation of systematic errors across large datasets. Coordinates should be maintained in WGS84 until final visualization steps.

Export Formats: Quality-filtered data are exported to CSV or GeoJSON formats for compatibility with standard analysis and GIS tools. Processing statistics track raw pixel counts, retention percentages at each filtering stage, and final valid pixels for complete quality documentation.

3.2. Meteorological Data Integration

ERA5 atmospheric reanalysis (0.25° resolution, hourly frequency) provides critical meteorological context, particularly Planetary Boundary Layer Height (PBLH) essential for column-to-surface concentration conversion[24]. PBLH determines the atmospheric volume available for pollutant dispersion and is derived from ERA5 atmospheric profiles using bulk Richardson number criterion.

Temporal Alignment: TROPOMI observations cluster around 13:30 local time (± 15 minutes daily), while ERA5 provides hourly data. For observations within ± 30 minutes of ERA5 reference time, PBLH values are assigned directly. For greater time differences, linear interpolation between adjacent hourly values provides reasonable estimates.

Spatial Alignment: For each satellite pixel, the nearest ERA5 grid point is identified using efficient spatial indexing. This recognizes that ERA5 represents average conditions over ~ 31 km areas while TROPOMI provides point measurements at ~ 5.5 km resolution.

Additional Data: Integration with Copernicus Global Land Cover (100m) enables source classification. SRTM Digital Elevation Model (30m) provides topographic context. WorldPop density grids (100m) support exposure assessment for health impact studies.

3.3. Analytical Processing

Spatial Aggregation: Satellite pixels aggregate to regular geographic grids (0.1°, 0.25°, or 0.5° resolution). For each grid cell: mean concentration (primary statistic), median (robust measure), 95th percentile (pollution episodes), pixel count (coverage documentation), and standard deviation (variability quantification).

Temporal Aggregation: Observations aggregate to daily, weekly, monthly, seasonal, and annual summaries reducing noise, identifying seasonal/interannual variations, assessing long-term trends, and aligning with epidemiological study periods.

Trend Analysis: Non-parametric Mann-Kendall testing quantifies temporal pollution trends with statistical significance testing. Sen's slope provides robust trend magnitude estimates. Supports policy effectiveness evaluation.

Hotspot Detection: Persistent pollution hotspots identified as grid cells exceeding the 90th percentile of regional distribution. Percentile ranking enables prioritization of intervention efforts.

3.4. Quality Validation and Assessment

Temporal Coverage: Calendar days with valid observations are counted (target: 360-366 annually). Missing dates are classified by cause (cloud/corruption, orbital mechanics, processing failures). Mean observations per day assessed for temporal representation quality.

Spatial Coverage: Data representativeness confirmed by verifying pixel coordinate bounds, counting observations in subregions to identify geographic bias, and calculating spatial completeness (target: $>90\%$ grid cell coverage).

Statistical Characterization: Descriptive statistics (mean, standard deviation, range, quartiles) characterize distributions. Distribution shape assessment (skewness, kurtosis) identifies deviations from normality. Outlier detection flags values $>5\sigma$ from climatological mean.

Ground-Truth Validation: Where available, satellite estimates collocate with ground-based stations. Standard metrics include R^2 (explained variance), RMSE and MAE (estimation errors), bias (systematic deviation), relative bias (percentage systematic error). Typical validation $R^2 = 0.6-0.8$ with $RMSE = 5-15 \mu g/m^3$ depending on pollutant and region.

4. Discussion

Advantages: Complete automation enables multi-year, multi-pollutant, multi-region processing with minimal intervention. Open-source implementation ensures reproducibility and peer review capability. Open-access data and software eliminate financial barriers, democratizing satellite remote sensing access.

Limitations: Output quality depends on TROPOMI retrieval accuracy varying with atmospheric conditions and surface properties. Regions with persistent cloud cover, high aerosol loading, or bright surfaces experience reduced availability and increased uncertainty. Validation requires collocated ground measurements, sparse in data-poor regions where satellite monitoring is most valuable. Multi-year global processing requires substantial storage (terabytes) and computational time (days-weeks).

Integration: Processed data serve as training inputs for machine learning models predicting surface air quality[33][34], enable chemical transport model evaluation, support policy impact assessment through trend analysis, and facilitate public awareness via accessible visualization applications.

Future Development: Real-time processing of NRT products enables operational forecasting (2-3 hour latency). Machine learning integration improves satellite-to-surface conversion accuracy. Cloud deployment (Docker, AWS, Google Cloud) enables on-demand scalable processing. Extension to additional pollutants (particulate matter) and complementary sensors (Sentinel-3, MODIS) provides comprehensive monitoring.

5. Conclusion

This paper presented a comprehensive data pipeline for Sentinel-5P atmospheric monitoring addressing critical barriers to effective satellite remote sensing utilization. Through systematic automation, rigorous quality assurance, auxiliary data integration, and robust validation, the pipeline transforms complex satellite observations into actionable environmental intelligence accessible to the global research community.

The modular architecture, open-source implementation, and detailed documentation enable researchers across diverse institutions to leverage TROPOMI data for air quality assessment, emission monitoring, atmospheric chemistry research, and policy-relevant environmental analysis. Validation across multiple Indian metropolitan regions demonstrates robust performance suitable for operational monitoring.

As satellite capabilities advance with Sentinel-5, Sentinel-4, TEMPO, and GEMS, standardized processing frameworks become critical for ensuring reproducibility, enabling method comparison, and maximizing scientific return on Earth observation infrastructure investment. This pipeline provides a foundation for standardization while maintaining flexibility for customization.

By democratizing access to satellite-based atmospheric monitoring through open-access data and open-source tools, this work contributes to more equitable global air quality knowledge, potentially informing interventions that reduce the enormous public health burden of air pollution worldwide.

Compliance with ethical standards

Acknowledgments

The authors sincerely thank the Institute of Science, Nagpur, and its director for providing the necessary resources and support to conduct this research. Appreciation is also extended to dataspace.copernicus.eu for their timely assistance.

Disclosure of conflict of interest

The authors declare that there is no conflict of interest.

References

- [1] Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., & Levelt, P. F. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, 120, 70-83.
- [2] Sentinel Online. (2024). Sentinel-5P TROPOMI User Guide. European Space Agency. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-5p-tropomi>

- [3] World Health Organization. (2023). WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. Geneva: World Health Organization.
- [4] Health Effects Institute. (2025). State of Global Air 2025: Special Report on Air Quality Monitoring Gaps. Boston, MA: Health Effects Institute.
- [5] Martin, R. V., Brauer, M., van Donkelaar, A., Shaddick, G., Narain, U., & Dey, S. (2019). No one knows which city has the highest concentration of fine particulate matter. *Atmospheric Environment: X*, 3, 100040.
- [6] Eskes, H., van Geffen, J., Boersma, K. F., & Eichmann, K. U. (2023). Sentinel-5 Precursor/TROPOMI Level 2 Product User Manual Nitrogen Dioxide. Royal Netherlands Meteorological Institute (KNMI). Document version 4.1.0.
- [7] Cooper, M. J., Martin, R. V., Hammer, M. S., Levelt, P. F., Veefkind, P., Lamsal, L. N., Krotkov, N. A., Brook, J. R., & McLinden, C. A. (2020). Global fine-scale changes in ambient NO₂ during COVID-19 lockdowns. *Nature*, 601(7891), 380-387.
- [8] Griffin, D., Zhao, X., McLinden, C. A., Boersma, F., Bourassa, A., Dammers, E., Degenstein, D., Eskes, H., Fehr, L., Fioletov, V., Hayden, K., Kharol, S. K., Li, S. M., Makar, P., Martin, R. V., Mihele, C., Mittermeier, R. L., Krotkov, N., Sneep, M., Lamsal, L. N., ter Linden, M., van Geffen, J., Veefkind, P., & Wolde, M. (2019). High-resolution mapping of nitrogen dioxide with TROPOMI: First results and validation. *Atmospheric Measurement Techniques*, 12(2), 1037-1053.
- [9] Liu, F., Beirle, S., Zhang, Q., van der A, R. J., Zheng, B., Tong, D., & He, K. (2019). NO_x emission trends over Chinese cities estimated from OMI observations during 2005-2015. *Atmospheric Chemistry and Physics*, 17(15), 9261-9275.
- [10] Goldberg, D. L., Anenberg, S. C., Griffin, D., McLinden, C. A., Lu, Z., & Streets, D. G. (2020). Disentangling the impact of the COVID-19 lockdowns on urban NO₂ from natural variability. *Geophysical Research Letters*, 47(17), e2020GL089269.
- [11] Lorente, A., Boersma, K. F., Eskes, H. J., Veefkind, J. P., van Geffen, J. H., de Zeeuw, M. B., van der Gon, H. A. C. D., Beirle, S., & Krol, M. C. (2019). Quantification of nitrogen oxides emissions from build-up of pollution over Paris with TROPOMI. *Scientific Reports*, 9, 20033.
- [12] Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., & Wilson, P. (2014). Best practices for scientific computing. *PLOS Biology*, 12(1), e1001745.
- [13] Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLOS Computational Biology*, 9(10), e1003285.
- [14] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362.
- [15] European Space Agency. (2024). Copernicus Open Access Hub: API Guide. ESA Copernicus Programme Documentation.
- [16] Hardt, D. (2012). The OAuth 2.0 Authorization Framework. RFC 6749. Internet Engineering Task Force.
- [17] OASIS. (2014). OData Version 4.0 Part 1: Protocol. OASIS Standard. Organization for the Advancement of Structured Information Standards.
- [18] McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 56-61.
- [19] Rew, R., & Davis, G. (1990). NetCDF: An interface for scientific data access. *IEEE Computer Graphics and Applications*, 10(4), 76-82.
- [20] van Geffen, J., Eskes, H., Boersma, K. F., & Maasakkers, J. D. (2022). TROPOMI ATBD of the total and tropospheric NO₂ data products. Royal Netherlands Meteorological Institute (KNMI). Document S5P-KNMI-L2-0005-RP, Issue 2.4.0.
- [21] Boersma, K. F., Eskes, H. J., Richter, A., De Smedt, I., Lorente, A., Beirle, S., van Geffen, J. H. G. M., Zara, M., Peters, E., Van Roozendaal, M., Wagner, T., Maasakkers, J. D., van der A, R. J., Nightingale, J., De Rudder, A., Irie, H., Pinardi, G., Lambert, J. C., & Compennolle, S. C. (2018). Improving algorithms and uncertainty estimates for

satellite NO₂ retrievals: Results from the quality assurance for the essential climate variables (QA4ECV) project. *Atmospheric Measurement Techniques*, 11(12), 6651-6678.

- [22] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., & Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999-2049.
- [23] Lamsal, L. N., Krotkov, N. A., Vasilkov, A., Marchenko, S., Qin, W., Yang, E. S., Fasnacht, Z., Joiner, J., Choi, S., Haffner, D., Swartz, W. H., Fisher, B., & Bucsela, E. (2021). Ozone Monitoring Instrument (OMI) Aura nitrogen dioxide standard product version 4.0 with improved surface and cloud treatments. *Atmospheric Measurement Techniques*, 14(1), 455-479.
- [24] Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X. L., Choi, D., Cheang, W. K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., & Frederick, J. E. (1998). Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *Journal of Geophysical Research: Atmospheres*, 103(D14), 17149-17161.
- [25] de Foy, B., Lu, Z., Streets, D. G., Lamsal, L. N., & Duncan, B. N. (2015). Estimates of power plant NO_x emissions and lifetimes from OMI NO₂ satellite retrievals. *Atmospheric Environment*, 116, 1-11.
- [26] Duncan, B. N., Lamsal, L. N., Thompson, A. M., Yoshida, Y., Lu, Z., Streets, D. G., Hurwitz, M. M., & Pickering, K. E. (2016). A space-based, high-resolution view of notable changes in urban NO_x pollution around the world (2005-2014). *Journal of Geophysical Research: Atmospheres*, 121(2), 976-996.
- [27] Ialongo, I., Virta, H., Eskes, H., Hovila, J., & Douros, J. (2020). Comparison of TROPOMI/Sentinel-5 Precursor NO₂ observations with ground-based measurements in Helsinki. *Atmospheric Measurement Techniques*, 13(1), 205-218.
- [28] Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillerá-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913-929.
- [29] Central Pollution Control Board. (2024). National Air Quality Monitoring Programme. Ministry of Environment, Forest and Climate Change, Government of India. <https://cpcb.nic.in/>
- [30] 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., & Williams, D. (2020). 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(11), 6113-6140.
- [31] Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M. B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L. J., & Schwartz, J. (2019). An ensemble-based model of PM_{2.5} concentration across the contiguous United States with high spatiotemporal resolution. *Environment International*, 130, 104909.
- [32] Zhan, Y., Luo, Y., Deng, X., Grieneisen, M. L., Zhang, M., & Di, B. (2018). Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. *Environmental Pollution*, 233, 464-473.