AI4O3: A Foundational Data Collection for Artificial Intelligence in Tropospheric Ozone Research

Makoto Kelp
University of Utah

Sebastian Hickman
European Centre for Medium-Range Weather Forecasts (ECMWF)

Kazuyuki Miyazaki
Jet Propulsion Laboratory (JPL)
California Institute of Technology

Kai-Lan Chang
Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder

Paul Griffiths
Bristol University

Qindan Zhu
Harvard–Smithsonian Center for Astrophysics

Gerbrand Koren
Utrecht University

Fernando Iglesias-Suárez
Predictia Intelligent Data Solutions S.L., Spain

Elyse Pennington
Jet Propulsion Laboratory (JPL)
California Institute of Technology

Martin Schultz
Forschungszentrum Jülich, Jülich Supercomputing Centre, University of Cologne

Scientific importance of the task. Ozone (O_3) is an internationally regulated air pollutant that harms human health[1,2], damages vegetation[3], and acts as a short-lived climate forcer in the upper atmosphere[4]. Beyond these impacts, **predicting tropospheric ozone with high accuracy expresses the ultimate skill of climate and atmospheric forecasting models**. Ozone is challenging to predict, both for physical and AI models, due to its coupling with nearly all processes in the atmosphere. It is photochemically produced in the presence of sunlight through reactions involving precursor gases emitted both from natural and anthropogenic sources[5,6]. Ozone can be transported on urban to intercontinental scales, and its lifetime can range from days to weeks, depending on environmental conditions[7]. Moreover, ozone frequently co-occurs with extreme events including wildfires and heatwaves[8–11].

Despite these challenges, ozone is in fact the most measured trace gas in our observational history, with reliable observations since the mid-20$^{\text{th}}$ century. In the satellite era, tropospheric ozone concentrations have shown a global increase, yet climate models often disagree on the spatial distribution and magnitude of this increase[12]. This discrepancy remains one of the largest open mysteries in atmospheric science modeling and may be due to uncertainties in emissions (tropical, traffic, soils)[13], nonlinear or missing chemistry[14,15], stratosphere-troposphere exchange[16], boundary-layer mixing[17], and deposition[18]. As a result, ozone is often oversimplified or even excluded altogether from climate and AI-based atmosphere models. Improving ozone prediction and diagnosing why models struggle will reveal how well these models capture interconnected Earth system processes, making ozone a powerful diagnostic for exposing model limitations and identifying processes that must be improved to build better physical and AI models of the climate system. To address these challenges, we propose a foundational benchmark dataset for tropospheric ozone: **AI4O3**.

What scientific question will the dataset enable? This training and benchmark dataset will enable the integration of tropospheric ozone prediction into the AI domain, which is currently dominated by weather forecasting and climate model emulation. Air quality research lacks well-defined benchmark datasets, limiting the application of AI in this realm[19,20]. Well-defined benchmarks, including datasets, training objectives, and evaluation scores, have been instrumental in advancing AI-driven weather forecasting between 2022 and 2024 (e.g., WeatherBench and WeatherBench2)[21,22]. Our benchmark ozone dataset will enable robust comparison between methods and guide the development of more accurate AI and physical models. Currently, less attention has been given to atmospheric composition datasets, where observations are often sparser and noisier than weather data. Microsoft's Aurora model is the only example of a foundational model fine-tuned for ozone prediction[23]. It uses data from CAMS reanalysis but lacks direct in-situ observations and provides predictions on 12 h time scales, which is too coarse for operational forecasts that impact human health and agriculture. Furthermore, most ozone reanalysis datasets (CAMS, MERRA-2) are coarse ($\sim$0.5–1°) and biased at the surface due to a lack of observational constraints.

A major challenge in ozone forecasting, whether using AI or physical models, lies in predicting extreme events. Extreme ozone episodes are low-probability, high-impact events that challenge model performance and uncertainty quantification. These extremes play a key role in assessing model uncertainty, helping to distinguish between epistemic and aleatoric errors. Evaluating AI models under extreme conditions, such as heatwaves and wildfires that are often linked with high ozone levels, provides a more rigorous evaluation of model performance and accuracy. AI models generally perform less accurately when predicting ozone extremes[23,24], emphasizing the need for a focused benchmark dataset. AI4O3 will enhance data quality for surrogate simulators of physical processes, enable fine-tuning of existing end-to-end LxMs, and advance the AI predictive capability for ozone-coupled extreme events.

Dataset rationale and strategy. Training and evaluating AI models requires high-quality, diverse data sources that are currently decentralized and challenging for traditional AI applications. Despite their availability, these data present challenges due to differences in data types, resolutions, and formats. While datasets such as ozone surface observations and CAMS reanalysis offer vast amounts of information, they remain barriers to AI model development due to messy, heterogeneous data formats and insufficient documentation focused on AI use. (See Appendix for acronyms and detailed data information presented below).

Our team includes experts in ozone observations, data assimilation, and AI; all required data and compute resources are available through public archives and collaborating institutions to create a comprehensive training and benchmark dataset. These sources (non-exhaustive) include the TOAR surface ozone database[26], the GHOST surface composition dataset[27], satellite observations (e.g., TROPOMI[28], TEMPO[29], GOME-2[30], TROPESS[31-32]), ozonesonde and LiDAR profiles[33–35], field campaigns (e.g., ATom[36], KORUS-AQ[37]), physical model outputs (e.g., GEOS-CF[38], GEOS-Chem[39]), reanalysis data (e.g., CAMS[40], MERRA-2[41,42]), and commercial aircraft measurements (e.g., IAGOS[43]). The varying datasets capture different aspects of ozone, illustrating the need for a harmonized benchmark. For example, surface observations provide hourly point measurements but are limited to station locations. In contrast, satellite data have kilometer-scale footprints and add information on spatial patterns with global coverage but limited sensitivity to the surface. Complementing these are ozonesondes, which provide vertical profiles but are spatially sparse, as are field campaign and commercial aircraft data. Physical models and reanalysis provide full spatial coverage at medium resolution ($\sim$0.5°) but inherit the biases of the physical models used. By combining these datasets into AI-ready formats, we aim to create the most comprehensive dataset characterizing ozone throughout the troposphere, providing a foundation for training large-scale, state-of-the-art AI models.

Metadata and covariates. We will include geolocation (latitude, longitude, altitude), time (timestamps, windows), data source (satellite, ozonesonde, model, etc.), quality flags, uncertainty estimates, and atmospheric state (e.g., temperature, pressure, relative humidity), plus complementary variables relevant to ozone formation (precursor emissions, NO_x, VOCs). These data are publicly available from multiple sources and do not require additional procurement by the team.

Curation, integration, and deployment. We will (1) provide cloud-optimized, AI-ready formats (e.g., Zarr, Icechunk) and document observations at native resolutions, spanning the 1970s to today; and (2) apply data assimilation[44,45] for AI to combine, regrid, and gap-fill on a global grid (4 km near the surface; 25 km in the free troposphere) with vertical levels every 25 hPa from the surface to the lower stratosphere, hourly from 2000 onward. We will host the data on Google Earth Engine[46] and leverage the scalable Ristretto library, which enables low-rank reconstructions of global atmospheric chemistry data using $\sim$1% of the original storage[47]. To start, we will release AI4O3v1 for North

America and Europe, where observations are densest, generated by an automated pipeline that ingests the large volume of ozone observations (see Appendix), standardizes units and metrics (e.g., parts per billion; maximum daily 8-hour average), and co-locates all measurements on a common hourly grid. Data will also be mirrored to a secure HPC-hosted repository, colocated with a major compute system to test and optimize dataset structure and training performance. We will reuse machine-readable ozone databases (e.g., TOAR) to avoid unnecessary reprocessing of established datasets. This framework is extensible, enabling easy incorporation of new observations and cost-effective storage, and delivering a faster ozone reanalysis product with improved surface estimates.

Acceleration and impact potential. AI4O3 will accelerate AI model development by providing a comprehensive, high-quality resource for training and validation. By enabling direct comparisons between AI and physical models, AI4O3 will guide more accurate ozone forecasting and benefit the next generation of weather and climate AI, given ozone's coupling to underlying processes. The resulting models will enhance operational forecasting, particularly for ozone and related extreme events (e.g., wildfires, heatwaves), supporting public health, climate resilience, and crop protection.

References

1. Anenberg, S. C. et al. Estimates of the Global Burden of Ambient $PM_{2.5}$, Ozone, and NO_2 on Asthma Incidence and Emergency Room Visits. *Environmental Health Perspectives* **126**, 107004 (2018).
2. Fleming, Z. L. et al. Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health. *Elem Sci Anth* **6**, 12 (2018).
3. Lefohn, A. S. et al. Tropospheric ozone assessment report: Global ozone metrics for climate change, human health, and crop/ecosystem research. *Elem Sci Anth* **6**, 27 (2018).
4. Monks, P. S. et al. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmospheric Chemistry and Physics* **15**, 8889–8973 (2015).
5. Archibald, A. T. et al. Tropospheric Ozone Assessment Report: Critical review of changes in the tropospheric ozone burden and budget from 1960–2100. *Elementa: Science of the Anthropocene* **8**, 034 (2020).
6. Lelieveld, J. & Dentener, F. J. What controls tropospheric ozone? *Journal of Geophysical Research: Atmospheres* **105**, 3531–3551 (2000).
7. Fiore, A. M. et al. Multimodel estimates of intercontinental source-receptor relationships for ozone pollution. *Journal of Geophysical Research* **114**, D04301 (2009).
8. Jaffe, D. A. & Wigder, N. L. Ozone production from wildfires: A critical review. *Atmospheric Environment* **51**, 1–10 (2012).
9. Schnell, J. L. & Prather, M. J. Co-occurrence of extremes in surface ozone, particulate matter, and temperature over eastern North America. *Proceedings of the National Academy of Sciences* **114**, 2854–2859 (2017).
10. Chang, K.-L., McDonald, B. C., Harkins, C. & Cooper, O. R. Surface ozone trend variability across the United States and the impact of heat waves (1990–2023). *Atmospheric Chemistry and Physics* **25**, 5101–5132 (2025).
11. Cooper, O. R. et al. Early Season 2023 Wildfires Generated Record-Breaking Surface Ozone Anomalies Across the U.S. Upper Midwest. *Geophysical Research Letters* **51**, e2024GL111481 (2024).
12. Christiansen, A., Mickley, L. J., Liu, J., Oman, L. D. & Hu, L. Multidecadal increases in global tropospheric ozone derived from ozonesonde and surface site observations: can models reproduce ozone trends? *Atmospheric Chemistry and Physics* **22**, 14751–14782 (2022).
13. Zhang, Y. et al. Contributions of World Regions to the Global Tropospheric Ozone Burden Change From 1980 to 2010. *Geophysical Research Letters* **48**, e2020GL089184 (2021).
14. Shah, V. et al. Nitrogen oxides in the free troposphere: implications for tropospheric oxidants and the interpretation of satellite NO_2 measurements. *Atmospheric Chemistry and Physics* **23**, 1227–1257 (2023).
15. Wang, S. et al. Active and widespread halogen chemistry in the tropical and subtropical free troposphere. *Proceedings of the National Academy of Sciences* **112**, 9281–9286 (2015).
16. Neu, J. L. et al. Tropospheric ozone variations governed by changes in stratospheric circulation. *Nature Geoscience* **7**, 340–344 (2014).
17. Lu, X. et al. Surface and tropospheric ozone trends in the Southern Hemisphere since 1990: possible linkages to poleward expansion of the Hadley circulation. *Science Bulletin* **64**, 400–409 (2019).

18. Clifton, O. E. et al. Dry Deposition of Ozone Over Land: Processes, Measurement, and Modeling. *Reviews of Geophysics* **58**, e2019RG000670 (2020).
19. Betancourt, C., Stomberg, T., Roscher, R., Schultz, M. G. & Stadtler, S. AQ-Bench: a benchmark dataset for machine learning on global air quality metrics. *Earth System Science Data* **13**, 3013–3033 (2021).
20. Dueben, P. D. et al. Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook. *Artificial Intelligence for the Earth Systems* **1**, e210002 (2022).
21. Rasp, S. et al. WeatherBench: a benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems* **12**, e2020MS002203 (2020).
22. Rasp, S. et al. WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems* **16**, e2023MS004019 (2024).
23. Bodnar, C. et al. A foundation model for the Earth system. *Nature* **641**, 1180–1187 (2025).
24. Hickman, S. H., Griffiths, P. T., Nowack, P. J. & Archibald, A. T. Short-term forecasting of ozone air pollution across Europe with transformers. *Environmental Data Science* **2**, e43 (2023).
25. Leufen, L. H., Kleinert, F. & Schultz, M. G. O3ResNet: A Deep Learning–Based Forecast System to Predict Local Ground-Level Daily Maximum 8-Hour Average Ozone in Rural and Suburban Environments. *Artificial Intelligence for the Earth Systems* **2** (2023).
26. Van Malderen, R. et al. Global ground-based tropospheric ozone measurements: reference data and individual site trends (2000–2022) from the TOAR-II/HEGIFTOM project. *Atmospheric Chemistry and Physics* **25**, 7187–7225 (2025).
27. Bowdalo, D. et al. GHOST: a globally harmonised dataset of surface atmospheric composition measurements. *Earth System Science Data* **16**, 4417–4495 (2024).
28. Garane, K. et al. TROPOMI/S5P total ozone column data: global ground-based validation and consistency with other satellite missions. *Atmospheric Measurement Techniques* **12**, 5263–5287 (2019).
29. Jin, X., Yang, Y., Gonzalez Abad, G., Nowlan, C. & Liu, X. Observing the Diurnal Variations of Ozone–NO_x–VOC Chemistry Over the U.S. From the Geostationary TEMPO Instrument. *Geophysical Research Letters* **52**, e2025GL116394 (2025).
30. Miles, G. M., Siddans, R., Kerridge, B. J., Latter, B. G. & Richards, N. A. D. Tropospheric ozone and ozone profiles retrieved from GOME-2 and their validation. *Atmospheric Measurement Techniques* **8**, 385–398 (2015).
31. Pennington, E. A. et al. Quantifying biases in TROPESS AIRS, CrIS, and joint AIRS+OMI tropospheric ozone products using ozonesondes. *Atmospheric Chemistry and Physics* **25**, 8533–8552 (2025).
32. Keppens, A. et al. Harmonisation of sixteen tropospheric ozone satellite data records. *EGUsphere* 1–38 (2025). doi:10.5194/egusphere-2024-3746.
33. Jiang, Y. B. et al. Validation of Aura Microwave Limb Sounder Ozone by ozonesonde and lidar measurements. *Journal of Geophysical Research: Atmospheres* **112** (2007).
34. Kuang, S. et al. Stratosphere-to-troposphere transport revealed by ground-based lidar and ozonesonde at a midlatitude site. *Journal of Geophysical Research: Atmospheres* **117** (2012).
35. Clain, G. et al. Tropospheric ozone climatology at two Southern Hemisphere tropical/subtropical sites (Reunion Island and Irene, South Africa) from ozonesondes, LIDAR, and in situ aircraft measurements. *Atmospheric Chemistry and Physics* **9**, 1723–1734 (2009).
36. Thompson, C. R. et al. The NASA Atmospheric Tomography (ATom) Mission: Imaging the Chemistry of the Global Atmosphere. (2022). doi:10.1175/BAMS-D-20-0315.1.
37. Schroeder, J. R. et al. Observation-based modeling of ozone chemistry in the Seoul metropolitan area during the Korea–United States Air Quality Study (KORUS-AQ). *Elementa: Science of the Anthropocene* **8**, 3 (2020).
38. Keller, C. A. et al. Description of the NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0. *Journal of Advances in Modeling Earth Systems* **13**, e2020MS002413 (2021).
39. Eastham, S. D. et al. GEOS-Chem High Performance (GCHP v11-02c): a next-generation implementation of the GEOS-Chem chemical transport model for massively parallel applications. *Geoscientific Model Development* **11**, 2941–2953 (2018).
40. Flemming, J. et al. The CAMS interim Reanalysis of Carbon Monoxide, Ozone and Aerosol for 2003–2015. *Atmospheric Chemistry and Physics* **17**, 1945–1983 (2017).

41. Wargan, K. et al. Evaluation of the Ozone Fields in NASA's MERRA-2 Reanalysis. *Journal of Climate* **30**, 2961–2988 (2017).

42. Wargan, K. et al. M2-SCREAM: A Stratospheric Composition Reanalysis of Aura MLS Data With MERRA-2 Transport. *Earth and Space Science* **10**, e2022EA002632 (2023).

43. Wang, H. et al. Global tropospheric ozone trends, attributions, and radiative impacts in 1995–2017: an integrated analysis using aircraft (IAGOS) observations, ozonesonde, and multi-decadal chemical model simulations. *Atmospheric Chemistry and Physics Discussions* 1–55 (2022). doi:10.5194/acp-2022-381.

44. Arroyo, A., Herrero, A., Tricio, V., Corchado, E. & Wozniak, M. Neural models for imputation of missing ozone data in air-quality datasets. *Complexity* 7238015 (2018). doi:10.1155/2018/7238015.

45. Betancourt, C. et al. Global, high-resolution mapping of tropospheric ozone – explainable machine learning and impact of uncertainties. *Geoscientific Model Development Discussions* 1–36 (2022). doi:10.5194/gmd-2022-2.

46. Gorelick, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **202**, 18–27 (2017).

47. Velegar, M., Erichson, N. B., Keller, C. A. & Kutz, J. N. Scalable diagnostics for global atmospheric chemistry using Ristretto library (version 1.0). *Geoscientific Model Development* **12**, 1525–1539 (2019).

48. NNJA Observations for Earth System Reanalysis: NOAA Physical Sciences Laboratory. `https://psl.noaa.gov/data/nnja_obs/`.

Appendix: Data sources, scale & resolution

A. Surface networks

- TOAR (Tropospheric Ozone Assessment Report) surface database
  Type: In situ surface ozone
  Spatial resolution: Point (station)
  Time resolution: Hourly (station dependent)
  Period: ~1970s–present, 6000 sites globally
- GHOST (Globally Harmonised Observations in Space and Time)
  Type: In situ surface ozone (and other chemical species)
  Spatial resolution: Point (station)
  Time resolution: Hourly to daily (network dependent)
  Period: ~1970s–present, over 7 billion measurements taken across 38 networks

B. Satellites

- TROPOMI (TROPOspheric Monitoring Instrument on Sentinel-5 Precursor)
  Type: Total and tropospheric column ozone; ozone profile product
  Spatial resolution: ~5.5×3.5 km at nadir (Level-2 column); profiles are coarser
  Time resolution: Daily global swath (polar orbiter)
  Period: 2018–present
- TEMPO (Tropospheric Emissions: Monitoring of Pollution instrument on Intelsat-40e)
  Type: Column ozone (and other chemical species); ozone profile product
  Spatial resolution: ~2.0×4.75 km (Level-2 column); profiles at ~8×4.75 km
  Time resolution: Hourly (daylight) over North America (geostationary)
  Period: 2023–present
- GOME-2 (Global Ozone Monitoring Experiment-2 on MetOp-A/B/C)
  Type: Total column ozone (UV-Vis)
  Spatial resolution: ~40×40 km (MetOp-A); ~80×40 km (MetOp-B/C)
  Time resolution: Near-daily global (polar orbiters)
  Period: 2006–present
- MLS (Microwave Limb Sounder on Aura)
  Type: Ozone profiles (focused on upper troposphere)
  Spatial resolution: ~6×165 km (cross-track × along-track) at tangent point; vertical resolution ~3 km near the tropopause

Time resolution: ~240 limb scans per orbit; near-daily global coverage (polar orbiter)
Period: 2004–present

- OMI (Ozone Monitoring Instrument on Aura)
  Type: Total column ozone (UV-Vis); other trace gases
  Spatial resolution: 13×24 km at nadir (Level-2 pixels)
  Time resolution: Daily global swath (polar orbiter)
  Period: 2004–present

- OMPS (Ozone Mapping and Profiler Suite on Suomi National Polar-orbiting Partnership (NPP), NOAA-20, NOAA-21)
  Type: Nadir Mapper (total column), Nadir Profiler (ozone profiles), Limb Profiler (ozone profiles; Suomi NPP only)
  Spatial resolution:
  Nadir Mapper: ~50×50 km at nadir on Suomi NPP; finer on newer platforms (~10×10 km on NOAA-21)
  Nadir Profiler: ~250×250 km at nadir on Suomi NPP; ~50×50 km on NOAA-20
  Limb Profiler: ~36×48 km horizontal; ~1 km vertical sampling (Suomi NPP)
  Time resolution: Nadir Mapper daily global; Nadir Profiler ~12-day global cycle; Limb Profiler daily limb tracks
  Period: 2011–present across the OMPS constellation with Suomi NPP from 2011–present, NOAA-20 2017–present, and NOAA-21 2022–present

- CrIS (Cross-track Infrared Sounder on Suomi NPP, NOAA-20, NOAA-21)
  Type: Ozone profiles from thermal IR retrievals
  Spatial resolution: ~14 km at nadir, coarser off-nadir
  Time resolution: Near-daily global swaths with ~2 passes per day (polar orbiters)
  Period: Suomi NPP 2011–present; NOAA-20 2017–present; NOAA-21 2022–present

- TROPESS (TRopospheric Ozone and its Precursors from Earth System Sounding; JPL retrieval suite using AIRS/CrIS/OMI)
  Type: Ozone profiles (thermal IR + UV-Vis)
  Spatial resolution: AIRS (Atmospheric Infrared Sounder) ~13.5 km; CrIS (Cross-track Infrared Sounder) ~14 km at nadir; OMI (Ozone Monitoring Instrument) ~13×24 km for column and profiles
  Time resolution: Near-daily global swaths (polar orbiters)
  Period: AIRS 2002–present, CrIS 2011–present, OMI 2004–present

C. Vertical profiles

- Ozonesondes
  Type: In situ ozone profiles (surface to ~30–35 km)
  Spatial resolution: Point (station), vertical ~100–150 m
  Time resolution: Usually weekly at long-term sites
  Period: ~1960s–present

- Ozone LiDAR: NDACC (Network for the Detection of Atmospheric Composition Change), TOLNet (Tropospheric Ozone LiDAR Network)
  Type: Remote-sensed ozone profiles (lower/mid troposphere)
  Spatial resolution: Point (station), vertical ~150–750 m
  Time resolution: ~10-minute retrievals
  Period: ~1990s–present

D. Field campaigns and aircraft

- ATom (Atmospheric Tomography Mission)
  Type: In situ ozone (and other chemical species) along global flight transects
  Spatial resolution: Along-track; 1–10 s sampling; vertical profiling ~0.2–12 km
  Time resolution: Per-flight, across four seasonal deployments
  Period: 2016–2018

- KORUS-AQ (Korea–United States Air Quality Study)
  Type: In situ and remote ozone, aircraft and surface measurements
  Spatial resolution: Regional (Korean peninsula/Seoul metro); along-track aircraft and site

stations
Time resolution: Intensive field campaign monitoring
Period: May–June 2016

- IAGOS (In-service Aircraft for a Global Observing System)
  Type: In situ ozone along commercial aircraft routes
  Spatial resolution: Along-track at cruise altitudes (~9–12 km)
  Time resolution: Continuous flights
  Period: 2011–present, with MOZAIC (Measurement of OZone by Airbus In-service air-Craft) predecessor data from 1994–2014

E. Models, reanalysis, and archives

- GEOS-CF (Goddard Earth Observing System–Composition Forecast)
  Type: Chemistry-coupled forecasts/analyses of ozone (and other chemical species)
  Spatial resolution: 0.25° global
  Time resolution: Hourly
  Period: 2018–present

- GEOS-Chem (Goddard Earth Observing System–Chemistry)
  Type: Simulated ozone and other chemical species from offline chemical transport model
  Spatial resolution: Global $4^\circ \times 5^\circ$ or $2^\circ \times 2.5^\circ$; nested $0.25^\circ \times 0.3125^\circ$ regions
  Time resolution: Typically hourly
  Period: User-defined simulations; usually 2000s to present

- CAMS Global Reanalysis–EAC4 (Copernicus Atmosphere Monitoring Service – ECMWF Atmospheric Composition Reanalysis 4)
  Type: Atmospheric composition reanalysis
  Spatial resolution: $0.75^\circ \times 0.75^\circ$ global
  Time resolution: 3-hourly
  Period: 2003–2024 with updates using a 12 h fixed reanalysis window

- MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2)
  Type: Atmospheric reanalysis including both ozone and meteorology fields
  Spatial resolution: $0.5^\circ \times 0.625^\circ$ global
  Time resolution: Hourly to 3-hourly
  Period: 1980–present

- NNJA (NOAA–NASA Joint Archive of Observations for Earth System Reanalysis)[48]
  Type: Atmospheric reanalysis and observation archive including ozone satellite observations from MLS, GOME-2, OMI, and OMPS
  Spatial resolution: Satellite resolutions above
  Time resolution: Hourly to daily
  Period: 1979–present