CHAPTER IV

Thesis Materials and Sources

Many surface O3 estimations have been created with similar air pollution measurements, meteorological imagery, and high-resolution estimations via combinations between them (Alvarez-Mendoza, Teodoro, and Ramirez-Cando 2019; Schultz, Schröder, Lyapina, Cooper, Galbally, Zhiqiang, et al. 2017a; W. Wang et al. 2022). Similar methods of kriging and complex trend estimation at high temporal resolutions (i.e. hourly, daily and weekly) are limited by the availability of imagery due to orbital patterns and geographic aspects of Earth. The National Oceanic and Atmospheric Administration/National Centers for Environmental Prediction (NOAA/NCEP), European Centre for Medium-Range Weather Forecasts (ECMWF), National Aeronautics and Space Administration/Global Modeling and Assimilation Office (NASA/GMAO), Japan Meteorological Agency (JMA), etc.; provide rich ensembles of climate data products which can be further combined in overlapping areas as a result of unified interests across nations e.g. (Carrillo-Torres, Hernández-Paniagua, and Mendoza 2017; XM Hu et al. 2022; Schultz, Schröder, Lyapina, Cooper, Galbally, Zhiqiang, et al. 2017a). This section details on datasets both used after filtering through the myriads of features depicted in Table VIII.1.3. and other materials available for future work and similar predictions. These sources all have individual data bases, many of which have been conveniently wrapped into a cloud based Big Data source hosted by Google called Earth Engine (GEE).

IV.1 Satellite Data

Most of this project ustilized raster data. As a representation of signals, raster data is the primary output for remote sensing and satellite technology. Remote sensing is a field dedicated to Earth-bound sun-sourced light while satellites While this type of data is frequently referred to as a “raster” in geography, it’s more simply a lattice of values; depicting whatever we want them to. If, these values come from a signal, then the proper transformations can be applied to properly represent its metaphorical value. For instance, temperature values can be estimated with extreme precision given the difference in Near Infrared (NIR) and incoming UV radiation. Much of this data can be seen on Google Earth Engine’s (GEE) vast cloud database. GEE was accessed to set the framework for the images used in the final predictive model. GEE is designed for planetary-scale environmental data analysis (Cardille et al. 2024) and provides access to a vast repository of satellite imagery and geospatial datasets, along with tools based in java for analysis and visualization [Site studies that used GEE].

IV.2. From Google Earth Engine to a Machine Near You

GEE is designed for planetary-scale environmental data analysis (Gorelick et al. 2017) and provides access to a vast repository of satellite imagery and geospatial datasets, along with tools based in java for analysis and visualization (Figure VIII.2.4.). As a Big Data source, this overwhelming repository of satellite imagery and geospatial datasets includes Landsat, Sentinel, MODIS, and more prominent satellite technologies. It’s being utilized more frequently in recent environmental monitoring, research, and management due to its accessibility. The cloud-based infrastructure of GEE allows for large-scale data processing without the need for local high-performance computing resources. This capability was particularly useful for SMaRK by processing and analyzing large datasets with precision and efficiency.

While the interactive code editor and Graphical User Interface (GUI), make it accessible to users well versed in JavaScript, it may be a difficult grasp for less code-savvy personnel. However, GEE enables collaborative work by allowing users to share scripts, data, and results easily for those willing to learn. The platform now has a strong Python integration provided by Dr. Wu at the University of Tennessee (Wu 2020). This package facilitated the automation of image gathering and is crucial to the reproducibility of this project. Feature integration into SMaRK was quintessential to this project and the many potential future uses among scientists, cooperate managers, and policymakers. GEE supports near-real-time data analysis, which is crucial for monitoring and responding to real-time environmental events. If translated from python to java correctly, SMaRK can work directly with GEE’s code editor, providing an essential air pollution model for the platform.

GEE provides free access to a vast archive of satellite imagery and geospatial datasets, including Landsat, Sentinel, MODIS, and more. This accessibility enables users to conduct historical analyses and display modern cycles with ease. Figure GEE depicts the code editor along with a quick representation of NDVI using Landsat 9. GEE has been utilized more frequently in recent environmental monitoring, research, and management due to its accessibility. The cloud-based infrastructure of GEE allows for large-scale data processing without the need for local high-performance computing resources. This capability is particularly useful for processing and analyzing large datasets quickly and efficiently.

While the interactive code editor and Graphical User Interface (GUI), make it accessible to users well versed in computer science, this data source may be a difficult grasp for less tech-savvy personnel. The platform has Python integration provided by Dr. Wu at the University of Tennessee. This package facilitated the automation of image gathering and is crucial to the reproducibility of this project. GEE enables collaborative work by allowing users to share scripts, data, and results easily. This feature integrated into SMaRK is quintessential to this project and potential future uses among scientists, cooperate managers, and policymakers. The platform supports near-real-time data analysis, which is crucial for monitoring and responding to real-time environmental events. When applied correctly, this project can work directly with GEE in JavaScript, providing an essential air pollution model for the platform.

IV.3. Sentinel-5p/TROPOMI Products

To address the necessity of improved data for analyzing environmental concerns, the Copernicus Atmosphere Monitoring Service (CAMS) has become one of the leading institutions in monitoring greenhouse gases, aiming at supporting policymakers, business and citizens with enhanced atmospheric environmental information (Guevara et al., 2021). The European Centre for Medium-Range Weather Forecasts (ECMWF), an independent intergovernmental organization supported by 35 states, implements CAMS behalf of the European Union for such purposes. CAMS launched the Sentinel-5 Precursor (S5P) in 2017, and it boasts the first Copernicus based mission dedicated to monitoring the planet’s atmosphere (Veefkind et al. 2012). The onboard sensor is frequently referred to as TROPOMI (TROPOspheric Monitoring Instrument). All of the S5P datasets, except CH4, have two versions: Near Real-Time (NRTI) and Offline (OFFL) except for CH4 (Methane) is available as OFFL only. The NRTI assets cover a smaller area than the OFFL assets but appear more quickly after acquisition. The OFFL assets contain data from a single orbit.

TROPOMI products are the main source of tropospheric ozone monitoring today, have been validated numerous times, and applied widely in prominent studies at global and national levels thanks to the relatively high accuracy, large geographic coverage, and long timespan (Zheng et al., 2019). This instrument offers a multitude of atmospheric variables including aerosol, NOx, and O­3 measurements which are used in this project. Only VOCs/pollutants which are known drivers of ozone and spatially significant were selected from the S5P satellite. Precursor aerosols such as SO2 concentrations tend to be localized and not representative of broad geographical areas such as in this study.

IV.3.1. AEROSOL INDEX

This dataset from the TROPOMI instrument provides high-resolution imagery of the UV Aerosol Index (UVAI) or the Absorbing Aerosol Index (AAI). Commonly referred to AAI, this index is based on wavelength-dependent changes in Rayleigh scattering seen in the UV spectrum for a specific pair of wavelengths, where AAI is the difference between observed and modelled reflectance [Zweers, 2024]. When an AAI value is positive, it indicates the presence of UV-absorbing aerosols like dust and smoke [De Graaf et al. 2005]. It is useful for tracking the evolution of episodic aerosol plumes from dust outbreaks, volcanic ash, and biomass burning which all have known correlations with ozone formation [Honrath et al., 2024]. The 354 nm and 388 nm wavelengths used have very low ozone absorption, so unlike aerosol optical thickness measurements, AAI can be calculated in the presence of clouds, allowing for daily global coverage. These rasters are available with daily averages from GEE and the CAMS home website at a resolution of 1113.2 meters. The collected AAI for the model represents a measure of the prevalence of aerosols in the atmosphere.

IV.3.2. NOX RETRIEVALS

Nitrogen oxides (NO2 and NO) are important trace gases in the Earth's atmosphere, present in both the troposphere and the stratosphere. They enter the atmosphere due to anthropogenic activities, such as fossil fuel combustion and biomass burning, and natural processes; such as wildfires, lightning, and microbiological processes in soils [Zheng et al., 2024]. The NO2 values represent concentrations of collective nitrogen oxides. This is to accurately portray the daytime photochemical cycle involving ozone (O3). This cycle converts NO into NO2 and vice versa on a timescale of minutes. The stratospheric vertical column of NO2 and NO2 slant column density bands were used as an estimation of the Ozone reaction happening in the atmosphere. This dataset is available at the same daily and 1113.2-meter resolution as the AAI.

IV.3.3. CARBON MONOXIDE RETRIEVALS

Carbon monoxide (CO) is an important atmospheric trace gas for understanding tropospheric chemistry. In certain urban areas, it is a major atmospheric pollutant. Main sources of CO are combustion of fossil fuels, biomass burning, and atmospheric oxidation of methane and other hydrocarbons. Whereas fossil fuel combustion is the main source of CO at northern mid-latitudes, the oxidation of isoprene and biomass burning play an important role in the tropics. TROPOMI on the Sentinel 5 Precursor (S5P) satellite observes the CO global abundance exploiting clear-sky and cloudy-sky Earth radiance measurements in the 2.3 μm spectral range of the shortwave infrared (SWIR) part of the solar spectrum. TROPOMI clear sky observations provide CO total columns with sensitivity to the tropospheric boundary layer. For cloudy atmospheres, the column sensitivity changes according to the light path.

IV.3.4. FORMALDEHYDE

Formaldehyde is an intermediate gas in almost all oxidation chains of non-methane volatile organic compounds (NMVOC), leading eventually to CO2. Non-Methane Volatile Organic Compounds (NMVOCs) are, together with NOx, CO and CH4, among the most important precursors of tropospheric O3. The major HCHO source in the remote atmosphere is CH4 oxidation. Over the continents, the oxidation of higher NMVOCs emitted from vegetation, fires, traffic and industrial sources results in important and localized enhancements of the HCHO levels. The seasonal and inter-annual variations of the formaldehyde distribution are principally related to temperature changes and fire events, but also to changes in anthropogenic activities. HCHO concentrations in the boundary layer can be directly related to the release of short-lived hydrocarbons, which mostly cannot be observed directly from space

IV.3.5. CLOUD MEASUREMENTS

Cloud properties are based on the OCRA and ROCINN algorithms currently being used on the GOME and GOME-2 products. OCRA retrieves cloud fraction measurements using UV and visible light spectral ranges while ROCINN retrieves the cloud height, pressure, and optical thickness (also called albedo). This is done by using measurements near the oxygen A-band at 760 nm. Version 3.0 of the algorithms are used, which are based on a more realistic treatment of clouds as optically uniform layers of light-scattering particles. Additionally, the cloud parameters are also provided for a cloud model which assumes the cloud to be a Lambertian reflecting boundary.

IV.3.6. NEAR-REAL TIME OZONE

For this TROPOMI product, there are two algorithms used to deliver total column ozone values: GDP for the near real-time and GODFIT for the offline products. This thesis focuses on the use of the GDP algorithm, as it is currently being used for generating the operational total ozone products from GOME, SCIAMACHY and GOME-2. These products have been used as the penultimate data in most atmospheric chemistry-based machine learning models as well as policies making for the EPA and EMWCF since 2017 [EPA, Cop, EMWCF, IPCC, 2017]. This product was used as a base line for ozone values and is available at the same resolution as the and NOx datasets.

IV.4. ERA5 COLLECTION

ERA5 refers to the fifth generation of European Re-Analysis (ERA), which is a climate reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) as part of the Copernicus Climate Change Service (C3S) [Dee et al. 2011; C3S, 2017]. ERA5 is the one of the latest climate reanalysis’ produced by ECMWF, providing hourly data on many atmospheric, land-surface and sea-state parameters together with estimates of uncertainty. The data is readily available in the Climate Data Store with grids at a 1km resolution, with atmospheric parameters on 37 pressure levels [Hersbach et al., 2020]. The dataset temporal estimates are from 1940 to the current day, with daily updates being made available 5 days behind real time. ERA5 has provided a consistent view of the evolution of land variables over several decades at an enhanced resolution to other land variable datasets such as the LMCS and GRID-Met dataset for meteorological data [Hersbach et al., 2020; Tarek et al., 2020; Muñoz-Sabater et al., 2021]. This reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using the laws of physics. As this data goes several decades back in time, it has been shown to provide accurate descriptions of climate change through many decades [Dee et al. 2011; Muñoz-Sabater et al., 2021]. This dataset includes 50 meteorological variables, of which, temperature, surface pressure, sum of total precipitation, short-wave radiation metrics, surface albedo, and northern/eastern wind velocities.

The asset used in this project is a daily aggregate of the ECMWF ERA5 Land hourly dataset on GEE, where daily aggregates have been pre-calculated to facilitate many applications requiring easy and fast access to the data. In GEE, the data is available from 1950 to three-months from real time [Muñoz-Sabater, 2019]. As meteorological variables are easier to model and monitor, this data is available at a coarse 11km. The resulting rasters were aggregated to a 500m resolution before being exported from GEE. This project utilized the 2m surface temperature, reflectivity of the Earth's surface, shortwave radiation at Earth’s surface, net change in solar radiation, Norther/Eastern components of wind up to a height of 10m, surface pressure, and total summation of precipitation as these were all suitable precursors for the ozone reaction [Gouw et al., 2024; Honrath et al., 2024; Telesca et al., 2024].

IV.5. VIIRS DAY/NIGHTTIME LIGHTS

The SNPP Visible Infrared Imaging Radiometer Suite (VIIRS) supports a Day-Night Band (DNB) sensor that provides global daily measurements of nocturnal visible and near-infrared (NIR) light that are suitable for Earth system science and applications. The VIIRS DNB's ultra-sensitivity in lowlight conditions enable us to generate a new set of science-quality nighttime products that manifest substantial improvements in sensor resolution and calibration when compared to the previous era of Defense Meteorological Satellite Program/Operational Linescan System's (DMSP/OLS) nighttime lights image products. Such improvements allow the VIIRS DNB products to better monitor both the magnitude and signature of nighttime phenomena, and anthropogenic sources of light emissions.

VNP46A2 is the short-name for the daily moonlight- and atmosphere-corrected Nighttime Lights (NTL) product called VIIRS/NPP Gap-Filled Lunar BRDF-Adjusted Nighttime Lights Daily L3 Global 500m Linear Lat Lon Grid.

The Suomi National Polar Orbiting Partnership (NPP), a near-polar geosynchronous orbit satellite with an orbital altitude of 824 km, carries a total of five sensors, and the Visible Infrared Imaging Radiometer Suite (VIIRS) is the most important of these. VIIRS has 22 spectral bands with a spectral range of “0.3∼14 μm”. Among them, the Day/Night Band (DNB) operates in the visible and near-infrared spectrum, between 500 and 900 nm, to collect low-light imaging data. This paper uses the NPP/VIIRS DNB dataset to characterize surface dynamics in the analysis of the impact factor of NO2 columns in China. Specifically, we used the DNB images from February 2018 to January 2019 and eliminated the image background noise.

IV.6. MODIS NDVI

MODIS vegetation indices, produced on 16-day intervals and at multiple spatial resolutions, provide consistent spatial and temporal comparisons of vegetation canopy greenness, a composite property of leaf area, chlorophyll and canopy structure. The main vegetation index is derived from atmospherically corrected reflectance in the red, near-infrared, and blue wavebands. These were used to produce a normalized difference vegetation index (NDVI). This provided continuity with NOAA's AVHRR NDVI time series record for historical and climate applications. Due to its ease of access and corrective model which minimizes canopy-soil variations and improves sensitivity over dense vegetation conditions, the resulting NDVI provided a highly accurate model of vegetation indices for use in the statistical models.

These vegetation indices are retrieved from daily, atmospherically corrected, bidirectional surface reflectance. The VI's use a MODIS-specific compositing method based on product quality assurance metrics to remove low quality pixels. From the remaining good quality VI values, a constrained view angle approach then selects a pixel to represent the compositing period (from the two highest NDVI values it selects the pixel that is closest-to-nadir). Because the MODIS sensors aboard Terra and Aqua satellites are identical, the VI algorithm generates each 16-day composite eight days apart (phased products) to permit a higher temporal resolution product by combining both data records. The MODIS VI product suite is now used successfully in all ecosystems, climate, and natural resources management studies and operational research as demonstrated by the ever-increasing body of peer publications. that each have commonalities with respect to spatial and spectral resolutions. The standard production run will process the NDVI/EVI at 250 m,500m,1km, and 0.05 Deg. resolution for 16-day and Monthly intervals. The output products will have data fields for the NDVI and EVI with corresponding QA, reflectance data, angular information and spatial statistics and std-dev of each VI and for the CMG scales. The 250m, 500m and 1km products are generated for each spatial tile (10 deg. x 10 deg. ~1200km x 1200km) in the sinusoidal projection. CMG products are generated globally in geographic projection.

IV.7. TOMS/OMI MERGED 10KM SURFACE OZONE PRODUCT

The Total Ozone Mapping Spectrometer (TOMS) data is produced and maintained by the Laboratory for Atmospheres at NASA's Goddard Space Flight Center. It yields a long-term, continuous record of the United States’ role in monitoring global and regional trends of surface ozone values over the past twenty-five years. The Ozone Monitoring Instrument (OMI) is an instrument aboard the Aura satellite and was launched more recently (July 2004 - current). OMI has a higher resolution (1.0 x 1.0 deg) and was used to scale the data observed from the TOMS instrument. Together, they represent a merged ozone product using observations from TOMS/EarthProbe, TOMS/Nimbus-7, TOMS/Meteor-3, OMI/Aura as well USGS based interpolations for dates with no data.

IV.8. EPA OZONE MONITOR DATA

This API offers access to a wide range of air quality data, including daily and hourly measurements of various pollutants other than ozone layers as well. Of the 65 available monitors in Arizona from 1980-2024, 50 monitors were found within Maricopa, Pinal, and Pima counties.

The importance of dependent variables is evaluated in Chapter for the final model. This was conducted mainly using shap plots, freely available in python (Lundberg et al., 2018, 2020). These features had multiple assignment variables associated with them to ensure each prediction had relevant satellite data per date it was available. For example, the Suomi National Polar-orbiting Partnership (SNPP) was launched Oct. 28, 2011, and is well established to offer statistical insights to the magnitude, signature, and anthropogenic sources of light emissions within the United States (Román et al., 2018). The Harmonized Global Nighttime Lights (HGNTL) dataset is available from 1992-2021 and conveniently implements calibrated DMSP-OLS NTL time series data for historical analysis of nighttime light imagery (Li et al., 2021).

The EPA offers data on numerous ground level pollutants and related metrological drivers (E. P. A. EPA 2025), providing exciting opportunities discussed later. Daily monitor data was collected starting from January 1st, 2018, to January 31st, 2025. The final date is noted for future work, making the potential stopping date the present day. The first max hour was utilized as exposure studies investigating ozone concentrations have found it to reflect the health impacts of long-term exposure better than short-term. (Hoek et al. 1997; Knowlton et al. 2004; Javanmardi et al. 2017; Anenberg et al. 2018; Anbari et al. 2022). The study area boundary was used to filter all monitor locations available in the state of Arizona for this time. This amounted to about 65 available monitors. Each monitor’s location and average maximum eight-hour surface ozone mixing ratio was extracted and averaged over the course of each available day within the yearly timeframe; a total of 5 years with 12 months for a total of 2,215 potential daily averages per monitoring site.

The date associated with each monitor was used in raster extraction as well, with Google Earth Engine (GEE) providing the rapid ability to assign date variable constraints (Hird et al. 2017; Cardille et al. 2024). There was a total of 50 monitors in the sample region, and a total of 30 complete monitors over the 5-year period. The distribution of training data can be seen in Figure X. This data was further processed to only include complete months (i.e all 30 or 31 days, leap years included) after feature creation. This was due to some of the features utilizing a smoothed weekly series described later in this thesis (Chapter 4; section IV). This led to some incomplete weeks near the tails of the dataset which were removed. 36 of the qualifying 50 monitoring locations were available in the PHOTUC region, with N = 61,344 samples avaliable in each model.

In python, monitor data was processed into maximum 8hr surface ozone values, location, and date. Geographic and temporal information were used to extract data from gathered imagery. Dummy temporal variables were implemented via OneHot encoding to give each model a temporal restriction. Month and seasonal values were implemented due to the known temporal patterns in existing literature (Shan et al. 2009; Honrath et al. 2017; S. Abdullah et al. 2019a; Balamurugan, Balamurugan, and Chen 2022). This same date was used as a filtering feature for the Residual Kriging method; however, it wasn’t used during the overall training and testing of statistical models. Implementing the date as a feature would add nearly 365 dummy features, increasing computation time and bias towards time in the models.

IV.9. ARIZONA BOUNDARIES

The Arizona state boundary was downloaded from TIGER line Census (United States Census Bureau 2022). The shapefiles and related files are an extract of select geographic information from the U.S. Census Bureau's Master Address File. A section within this vase database contains the Topologically Integrated Geographic Encoding and Referencing (TIGER) Database (MTDB). The MTDB represents a seamless national file with no overlaps or gaps between parts. They are an extremely convenient option for projects, as each data product can be further combined to cover the entire nation for later analysis.

The boundaries of incorporated places in the Arizona shapefile have been current since January 1, 2021, as reported through the US Census Bureau's Boundary and Annexation Survey (BAS) (US Census Bureau 2025). State-wide, county level data was downloaded, filtered and displayed [Figure X] to obtain the overall boundary of monitors to use. Each state- and county-level boundary file was run through a unifying process to remove inner state boundaries, leaving one singular shape of the study region containing Pheonix and Tuson, PHOTUC. The boundaries used as a reference to incorporate these two cities are Pinal, Pina, and Maricopa Counties.

IV.10. FINAL SURFACE OZONE DATASET

Each training feature was checked for correlation and normality prior to model training. Missing data in-between dates due to errors in satellite imagery (Cite Sat Error comments) was interpolated. Some values were extracted at 0 due to pre-processing malfunctions. These were imputed utilizing the KNN imputation with 6 nearest neighbors. The results of data interpolation and imputation can be seen in Table X. In addition, the effectiveness of the residual kriging method approach is tested on the Total Ozone Mapping Spectrometer (TOMS) dataset which has been combined with data from the Ozone Monitoring Instrument (OMI).

Citations

Abdullah, Samsuri, Najihah Husna Ahmad Nasir, Marzuki Ismail, Ali Najah Ahmed, and Mohamad Nor Khasbi Jarkoni. 2019. “Development of Ozone Prediction Model in Urban Area.” *International Journal of Innovative Technology and Exploring Engineering* 8 (10): 2263–67. doi:10.35940/ijitee.J1127.0881019.

Alvarez-Mendoza, CI, A Teodoro, and L Ramirez-Cando. 2019. “Spatial Estimation of Surface Ozone Concentrations in Quito Ecuador with Remote Sensing Data, Air Pollution Measurements and Meteorological Variables.” *ENVIRONMENTAL MONITORING AND ASSESSMENT* 191 (3). doi:10.1007/s10661-019-7286-6.

Anbari, Khatereh, Yusef Omidi Khaniabadi, Pierre Sicard, Hasan Raja Naqvi, and Rajab Rashidi. 2022. “Increased Tropospheric Ozone Levels as a Public Health Issue during COVID-19 Lockdown and Estimation the Related Pulmonary Diseases.” *Atmospheric Pollution Research* 13 (12): 101600. doi:10.1016/j.apr.2022.101600.

Anenberg, Susan C., Daven K. Henze, Veronica Tinney, Patrick L. Kinney, William Raich, Neal Fann, Chris S. Malley, et al. 2018. “Estimates of the Global Burden of Ambient PM2.5, Ozone, and NO2 on Asthma Incidence and Emergency Room Visits.” *Environmental Health Perspectives* 126 (10). Environmental Health Perspectives: 107004. doi:10.1289/EHP3766.

Balamurugan, Vigneshkumar, Vinothkumar Balamurugan, and Jia Chen. 2022. “Importance of Ozone Precursors Information in Modelling Urban Surface Ozone Variability Using Machine Learning Algorithm.” *Scientific Reports* 12 (1). Nature Publishing Group: 5646. doi:10.1038/s41598-022-09619-6.

Cardille, Jeffrey A., Morgan A. Crowley, David Saah, and Nicholas E. Clinton, eds. 2024. *Cloud-Based Remote Sensing with Google Earth Engine: Fundamentals and Applications*. Cham: Springer International Publishing. doi:10.1007/978-3-031-26588-4.

Carrillo-Torres, ER, IY Hernández-Paniagua, and A Mendoza. 2017. “Use of Combined Observational-and Model-Derived Photochemical Indicators to Assess the O3-NOx-VOC System Sensitivity in Urban Areas.” *ATMOSPHERE* 8 (2). doi:10.3390/atmos8020022.

EPA, Environmental Protection Agency. 2025. “Air Quality System (AQS) API.” Table. https://www.epa.gov/outdoor-air-quality-data/download-daily-data: Air Quality Data.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. “Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone.” *Remote Sensing of Environment* 202 (December). Elsevier BV: 18–27. doi:10.1016/j.rse.2017.06.031.

Hird, Jennifer, Evan DeLancey, Gregory McDermid, and Jahan Kariyeva. 2017. “Google Earth Engine, Open-Access Satellite Data, and Machine Learning in Support of Large-Area Probabilistic Wetland Mapping.” *Remote Sensing* 9 (12): 1315. doi:10.3390/rs9121315.

Hoek, Gerard, Joel D. Schwartz, Bernard Groot, and Paul Eilers. 1997. “Effects of Ambient Particulate Matter and Ozone on Daily Mortality in Rotterdam, the Netherlands.” *Archives of Environmental Health: An International Journal* 52 (6): 455–63. doi:10.1080/00039899709602224.

Honrath, Richard E., Yaoxian Huang, Shiliang Wu, Louisa J. Kramer, and Detlev Helmig. 2017. “Surface Ozone and Its Precursors at Summit, Greenland: Comparison between Observations and Model Simulations.” *Atmospheric Chemistry and Physics* 17 (23): 14661–74. doi:10.5194/acp-17-14661-2017.

Hu, XM, J Zhang, WH Xue, LH Zhou, YF Che, and T Han. 2022. “Estimation of the Near-Surface Ozone Concentration with Full Spatiotemporal Coverage across the Beijing-Tianjin-Hebei Region Based on Extreme Gradient Boosting Combined with a WRF-Chem Model.” *ATMOSPHERE* 13 (4). doi:10.3390/atmos13040632.

Javanmardi, Parviz, Pouran Morovati, Majid Farhadi, Sahar Geravandi, Yusef Omidi Khaniabadi, Kambiz Ahmadi Angali, Adewale Matthew Taiwo, et al. 2017. “Monitoring The Impact Of Ambient Ozone On Human Health Using Time Series Analysis And Air Quality Model Approaches.” *Fresenius Environmental Bulletin* 26 (November).

Knowlton, Kim, Joyce E. Rosenthal, Christian Hogrefe, Barry Lynn, Stuart Gaffin, Richard Goldberg, Cynthia Rosenzweig, Kevin Civerolo, Jia-Yeong Ku, and Patrick L. Kinney. 2004. “Assessing Ozone-Related Health Impacts under a Changing Climate.” *Environmental Health Perspectives* 112 (15): 1557–63. doi:10.1289/ehp.7163.

Schultz, Martin G., Sabine Schröder, Olga Lyapina, Owen R. Cooper, Ian Galbally, Irina Petropavlovskikh, Erika von Schneidemesser, et al. 2017. “Tropospheric Ozone Assessment Report: Database and Metrics Data of Global Surface Ozone Observations.” Edited by Michael E. Chang and Alastair Lewis. *Elementa: Science of the Anthropocene* 5 (October): 58. doi:10.1525/elementa.244.

Shan, Wenpo, Yongquan Yin, Haixia Lu, and Shuxuan Liang. 2009. “A Meteorological Analysis of Ozone Episodes Using HYSPLIT Model and Surface Data.” *Atmospheric Research* 93 (4): 767–76. doi:10.1016/j.atmosres.2009.03.007.

United States Census Bureau. 2022. “TIGER/Line Shapefiles (Machine Readable Data Files).” Department of Commerce. https://www.census.gov/cgi-bin/geo/shapefiles/index.php.

US Census Bureau. 2025. “Boundary and Annexation Survey (BAS).” *Census.Gov*. January 22. https://www.census.gov/programs-surveys/bas.html.

Wang, Wenhao, Xiong Liu, Jianzhao Bi, and Yang Liu. 2022. “A Machine Learning Model to Estimate Ground-Level Ozone Concentrations in California Using TROPOMI Data and High-Resolution Meteorology.” *Environment International* 158 (January): 106917. doi:10.1016/j.envint.2021.106917.

Wu, Qiusheng. 2020. “Geemap: A Python Package for Interactive Mapping with Google Earth Engine.” *Journal of Open Source Software* 5 (51): 2305. doi:10.21105/joss.02305.