Integration of Satellite, Models and Ground Sensors for City-scale Air Quality Monitoring through the Open Data Cube

Afshin Moazzam[1], Daniele Oxoli[1], Jesus Rodrigo Cedeno Jimenez[1], Maria Antonia Brovelli[1]

[1] Dept. of Civil and Environmental Engineering, Politecnico di Milano, P.zza Leonardo da Vinci 32, 20133, Milan, Italy - (afshin.moazzam, daniele.oxoli, jesusrodrigo.cedeno, maria.brovelli)@polimi.it

Keywords: Air Quality, Multi-sensors Data Integration, Open Data Cube, Smart Cities

Abstract

Effective air quality monitoring is crucial for assessing pollution levels and mitigating associated health risks, particularly in densely populated urban environments. Traditional ground-based monitoring stations provide accurate data but lack the spatial coverage necessary for comprehensive city-wide analysis. This study presents an integrated approach that combines satellite observations, atmospheric composition models, and ground sensor data within the Open Data Cube framework. The Open Data Cube system facilitates data access, processing, and analysis by structuring multi-source air quality information into a unified data endpoint. The Metropolitan City of Milan serves as a case study to evaluate this framework's potential. The results highlight the feasibility of this approach with the potential of enhancing operational air quality assessments to empower future monitoring services, including real-time applications and improved data accessibility for decision-making and public communication.

1. Introduction

Air pollution is recognized as the greatest environmental threat to global health and a critical sustainability challenge, closely linked to several United Nations Sustainable Development Goals (e.g., 3.9 and 11.6). Rising concentrations of common air pollutants, including Particulate Matter (PM), Ozone (O_3), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2), have been associated with increased hospital admissions. Overall, the deterioration of air quality is exacerbating the burden of respiratory and cardiovascular diseases and global mortality, even at low exposure levels (European Environment Agency, 2023).

In urban areas worldwide, air quality is a major concern due to high population density and the concentration of transportation and industrial activities, which together increase exposure levels and related health risks. While both natural and anthropogenic factors contribute to ambient air pollution, it is established that anthropogenic factors are primarily responsible for most air pollutant emissions. Given the severity of the issue, it is crucial to develop and empower comprehensive systems for continuous air quality monitoring, thereby providing decision-makers and regulators with accurate and up-to-date information on pollutant concentrations, emission sources, and spatiotemporal trends, to support the formulation of targeted policies and mitigation strategies (Chen et al., 2024). Conventionally, ground sensor measurements have been employed to monitor air quality. However, advancements in technology have led to the utilization of novel methods that leverage also non-conventional data from satellites and atmospheric composition models.

Ground measurement sensors provide reliable and direct assessments of pollutant concentrations at specific locations. These sensors enable the collection of high-precision data essential for regulatory compliance and local air quality assessments. Advancements in technology have introduced novel methods that complement traditional ground-based monitoring by integrating non-conventional data sources, such as satellite remote sensing and atmospheric composition models. Satellite observations offer extensive spatial coverage and enable the continuous tracking of air pollution patterns at both regional and global scales. Meanwhile, atmospheric composition models incorporate multiple data sources, including meteorological variables and emission inventories, to simulate pollutant dispersion and predict air quality trends. These approaches provide valuable insights into pollutant composition, transport mechanisms, and long-term trends, addressing gaps left by ground-based monitoring networks (Oxoli et al., 2024). The integration of multiple air quality data is promising for the development of a system capable of capturing various properties of air quality phenomena, wherein each component addresses a specific characteristic of the issue at different resolutions, thereby providing distinct information that can be synthesized to create a comprehensive view of the situation. Nevertheless, the integration of these diverse data sources presents significant challenges, necessitating the consideration of additional variables, such as meteorological observations, to enhance the reliability of air quality monitoring systems at both local and regional scales (Chen et al., 2024).

In view of the above, this study aims to contribute to the development of frameworks for the integration of multi-source air quality data, including satellite, atmospheric composition models, and ground-based sensors, to enhance air quality monitoring within future smart city governance practices. It is worth noticing that while the integration of databases with varying resolutions may yield complex information, the objective of this work is not to create a comparative database; rather, it is to establish a database that will be enriched by its inputs.

The proposed data management system for this integration is the free and open-source software Open Data Cube (ODC). The ODC eliminates the necessity for data pre-processing by the end user and mitigates the time and complexity associated with accessing gridded air pollution observations, along with any additional air quality-related variables (Cedeno Jimenez et al., 2021). By structuring multi-source air quality data into an efficient and scalable data handling framework, the ODC facilitates querying, visualization, and analysis, making it a powerful tool for researchers, policymakers, and environmental agencies.

An experimental implementation for the Metropolitan City of Milan (Northern Italy) and its surroundings is presented as a case study. This region, characterized by high population density and significant anthropogenic emissions, represents a relevant testbed for evaluating the effectiveness of integrating ground-based, satellite, and model-derived air pollution data within the ODC framework. The results from this study are expected to enhance traditional air quality monitoring practices by offering more comprehensive and near-real-time insights into pollutant distribution and trends.

2. Materials and Methods

2.1 Study Area

The Metropolitan City of Milan, located in the Lombardy region (Northern Italy), serves as the focal point for this research. Milan is located within the Po Valley (Figure 1), a region characterized by high urbanization and industrialization, as well as geomorphological conditions that hinder the dispersion of air pollutants (Raffaelli et al., 2020). As of 2024, the metropolitan area of Milan is home to approximately 3 million residents, making it the second most populous city in Italy, following Rome. Previous research has indicated that transportation and land use are among the primary contributors to air pollution, particularly through the exacerbation of traffic congestion in densely populated urban areas (Meroni et al., 2021) such as Milan. The large population and the necessity for an efficient public transportation system and urban design in this context heighten the risk of exposure to elevated pollution levels. Consequently, Milan serves as an exemplary case for the development and calibration of new monitoring initiatives. Given these characteristics, similar approaches can be implemented in other cities facing comparable conditions.

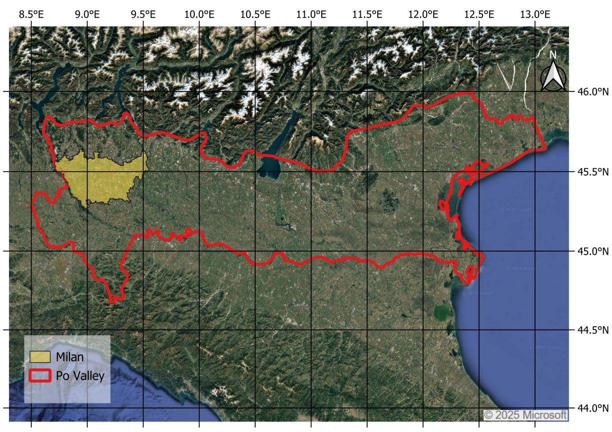

Figure 1. Map displaying the Metropolitan City of Milan within the Po Valley boundary.

2.2 Air Quality Data

The air pollutant species considered in this study include SO_2, O_3 carbon monoxide (CO), nitrogen oxides (NO_x), Particulate Matter ($PM_{2.5}$ and PM_{10}), and ammonia (NH_3), all of which are recognized as significant pollutants by the World Health Organization (World Health Organization, 2021). CO is known to be toxic to humans and can lead to a reduction in blood oxygen levels. NO_x is associated with the formation of haze, respiratory diseases, and a deterioration in air quality. SO_2 contributes to the formation of acidic gases that can adversely affect infrastructure and natural habitats. Although O_3 is a naturally occurring key component of the atmosphere, elevated concentrations can result in smog formation. PM is linked to lung cancer and cardiovascular diseases, while NH_3, at high concentrations, can induce respiratory and skin irritations (Sharma et al., 2013).

Meteorological variables encompass temperature, wind, precipitation, solar radiation, humidity, and surface pressure. These variables are considered in relation to the concentration and spatial distribution of pollutants (Zhang, 2019). However, they are not regarded as direct contributors to pollutant formation, a relationship that remains unexplored in the present study. The data analyzed pertain to the years 2022 and 2023, which were incorporated during the initial implementation of the system. The selection of satellite, model, and ground-sensor data sources prioritized the highest available temporal and spatial resolution, while also ensuring that the datasets were open-license and programmatically accessible from authoritative sources. The datasets of interest for this research include Sentinel-5P, Copernicus Atmospheric Monitoring System (CAMS), European Centre for Medium-Range Weather Forecasts (ECMWF) global atmospheric reanalysis (ERA-5-Land), and the Lombardy Region Environmental Protection Agency (ARPA) ground sensors network.

2.2.1 Satellite data: Sentinel-5P Sentinel 5P (S5P) satellite, launched on October 13, 2017, marks the first Copernicus mission dedicated to atmospheric monitoring. This mission is characterized as a single-satellite initiative that is equipped with the Tropospheric Monitoring Instrument (TROPOMI), an optical device capable of measuring ultraviolet, visible, shortwave infrared, and near-infrared spectral bands. This capability facilitates the measurement of various significant atmospheric gases. Due to its spatial resolution of 5.5 km by 3.5 km, effective since August 2019, S5P is particularly effective for studying air pollution in urban areas. The satellite operates in a sun-synchronous orbit at an altitude of 824 km above the Earth’s surface, and its 2,600 km-wide swath enables it to observe the entire planet on a daily basis. TROPOMI is oriented at a 30-degree angle to optimize its Earth observation capabilities. Data from S5P are disseminated to users at two levels: Level-1B (L1B), which includes geo-located, radiometrically corrected top-of-atmosphere data along with solar irradiances, and Level-2 (L2), which provides geophysical atmospheric parameters. The data are made available through two distinct methods based on their timing: Near Real-Time (NRT) data, accessible three hours post-sensing, and Non-Time Critical (NTC) data, which undergoes specific calibrations for enhanced accuracy and is available twelve hours after sensing (Bodah et al., 2022, Prunet et al., 2020, Virta et al., 2023, Zheng et al., 2019). L2 NTC validated data were selected and downloaded utilizing the Copernicus Application Programming Interface (API) (`https://documentation.dataspace.copernicus.eu/APIs/OData.html`) in Python. This data encompasses the total vertical columns of SO_2, O_3, and CO, as well as the tropospheric column of nitrogen dioxide NO_2. It is important to note that L2 data are organized according to the timing of the measurements and do not conform to a fixed spatial grid (Schneider et al., 2021). To enhance data management, further processing was conducted using the HARP toolbox (`https://atmospherictoolbox.org/harp`) to establish a standardized spatial grid. HARP is specifically designed for reading, processing, and inter-comparison of remotely sensed data. The data were resampled at a resolution of 1 km by 1 km, which has been deemed appropriate for the objectives of this study. This resampling process employs average weighting, whereby the toolkit generates bins based on the specified coordinates, and the values of the new bins correspond to the weighted averages of overlapping pixels from the L2 S5P data. Furthermore, HARP is applicable for additional processing steps relevant to our research. According to the documentation for each product, a quality assurance value has been established for each parameter to filter out low-quality pixels. This value is determined based on instrument performance, environmental conditions, and algorithm sensitivity. Lastly, HARP can be employed to blend acquisitions in overlapping regions. In instances where multiple images have been acquired for a given area of interest, HARP will produce a blended output based on the weighted averaging of those pixels. The code is designed to save the final files in NetCDF format, which is essential for integration into the ODC.

2.2.2 Atmospheric Composition Models: Copernicus Atmosphere Monitoring Service To accurately capture the variations within the Earth system, it is essential to consider all components of the system and their interactions. These components include the atmosphere, ocean, cryosphere, land, and biosphere. However, for specific objectives, it may be appropriate to focus on only a subset of these components when developing a model (Lahoz, 2003). General Circulation Models (GCMs) are particularly useful for this purpose, as they can simulate large-scale circulations across the entire planet over various time scales (Rafkin et al., 2001). GCMs can serve as the foundational framework for designing atmospheric models. On a large scale, the atmosphere can be regarded as being in a state of equilibrium; thus, any model should account for the fact that atmospheric conditions are typically in balance (Lauritzen et al., 2011). These models will manage the physical and chemical processes occurring in the atmosphere through the application of numerical techniques (Trini Castelli et al., 2017), which may be either linear or non-linear. The CAMS air quality forecast (`https://ads.atmosphere.copernicus.eu/datasets/cams-europe-air-quality-forecasts`) is an example of a model designed to monitor atmospheric composition at both global and regional scales.

The regional air quality production of the CAMS is of particular interest in this study. This production is based on the ensemble of eleven state-of-the-art numerical models (Copernicus Atmosphere Monitoring Service, 2025). The regional model of CAMS provides a four-day forecast and daily analyses for the European territory at ten altitude levels, ranging from 0 to 5000 meters above sea level. For the daily analyses of air pollutants, the system utilizes observations from one day prior and assimilates them with the models. The system is capable of integrating all models or a selection of models based on the requirements and requests of the user to produce the final product. In instances where all models are utilized, the median value of all individual outputs is calculated, which, according to current documentation, yields the most accurate results. The model inputs consist of meteorological forecasts from the ECMWF Integrated Forecasting System (IFS), while chemical boundary conditions are derived from the CAMS IFS-TM5 global production. Additionally, CAMS provides emission information (Douros et al., 2023).

CAMS regional data is provided at a spatial resolution of 0.1 by 0.1 degrees, with an hourly frequency. Data can be accessed through either an API or the website. For the purpose of simplicity and to facilitate automated access, we have chosen to utilize the ADS API from the cdsapi library in Python (`https://cds.climate.copernicus.eu/how-to-api`). The pollutants of interest include $PM2.5$, $PM10$, SO_2, NO_x, CO, O_3, and NH_3, which were retrieved at the surface level within the model's analysis mode. Although the CAMS system is limited in its ability to capture detailed information due to its relatively coarse spatial resolution, it has demonstrated significant effectiveness by integrating various databases, thereby playing a crucial role in pollution control and risk assessment (Casciaro et al., 2022). In this study, the capabilities of CAMS will be employed at its original resolution to monitor larger-scale trends.

2.2.3 Meteorological Models: ERA-5-Land Weather is often perceived as a transient phenomenon that influences short-term decision-making. However, it is equally important to examine the historical development and variability of weather patterns, as this knowledge can enhance our understanding of the relationship between weather and climate (Slivinski, 2018). One effective method for acquiring historical weather data is through atmospheric reanalyses. These reanalyses are generated using assimilation techniques that integrate observational data with numerical weather prediction models. In recent years, the significance of reanalyses has increased in the study of meteorological information. Although the results produced by reanalyses may closely resemble those obtained from direct observations and measurements, a key distinction lies in the fact that the uncertainties and errors associated with reanalyses are less well understood compared to those related to conventional measurements (Parker, 2016).

ERA-5 Land dataset (`https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land`) features a spatial resolution of 9 km, derived from downscaled meteorological forcing data from ERA-5, and is integrated with the land surface model numerically. Since it is computed under a single simulation and is not coupled with the atmospheric module of ECMFW IFS, it can be efficiently rerun and updated in a short timeframe if any modifications occur in the land component. The dataset provides information for 50 variables globally at an hourly frequency. As recommended by the provider at the time of writing this manuscript, the uncertainty fields corresponding to the ERA-5 dataset should be utilized as the uncertainty field for ERA-5 Land, as specific uncertainty fields for ERA-5 Land are not available. This limitation arises from the initial plan to present uncertainty fields based on a 10-member ensemble run; however, preliminary experiments indicated unrealistically low values due to the limited spread exhibited by the ERA-5 forcing fields (Dalla Torre et al., 2024, Gomis-Cebolla et al., 2023, Clelland et al., 2024, Copernicus Climate Change Service, 2025). Data can be accessed through the Climate Data Store (CDS) API, and a code has been developed to facilitate semi-automated retrieval. The retrieved components include temperature, wind, precipitation, and pressure. Through the CDS, it is possible to obtain the data in NetCDF format, which is suitable for integration into the ODC.

2.2.4 Ground-sensors: ARPA Ground sensors are utilized to collect measurements from the near-Earth surface. They are extensively employed for meteorological and air quality assessments. These sensors can be either fixed at a specific location or mobile, allowing for easy relocation. Compared to other measurement types, ground sensors exhibit greater reliability, as their uncertainty is more comprehensible and quantifiable. Furthermore, they can be integrated with various systems, such as models, as previously discussed in relation to these measurement types. Ground measurements play a crucial role in validation due to their accuracy and are frequently included in studies that rely on measurements from sources other than ground sensors.

For this research, the ARPA Lombardy sensor network (`https://www.arpalombardia.it/temi-ambientali/aria`) has been employed (Figure 2). ARPA operates in each Italian region with each agency responsible for its data processing and dissemination to users. The ARPA Lombardy network primarily consists of fixed sensors; however, it also includes a limited number of mobile sensors. The stations within this network are capable of measuring air quality, meteorological parameters, or both, and access to their data is publicly available (Maranzano, 2022). Users can either submit a request through the website to obtain specific parameters of interest from designated stations or utilize an API to generate the required data.

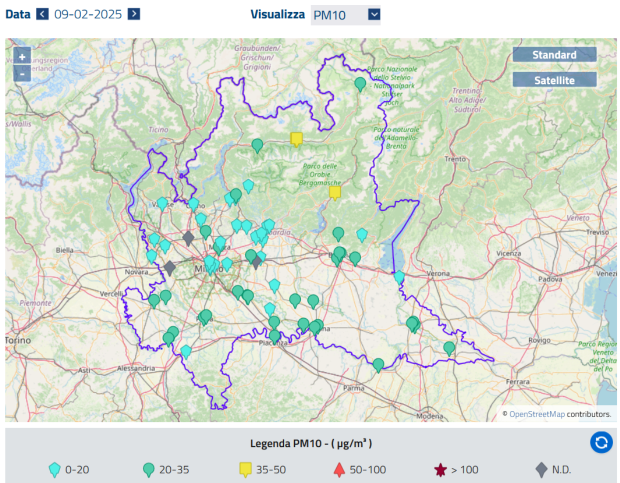

Figure 2. Locations of ARPA Lombardy air quality stations. Retrieved from: `https://www.arpalombardia.it/temi-ambientali/aria/stazioni-fisse`.

The data collected in this study from the ARPA includes $PM10$, $PM2.5$, SO_2, NO_2, CO, O_3, NH_3, temperature, wind, precipitation, solar radiation, and humidity. After downloading, the data are not immediately suitable for integration into the ODC and must be interpolated onto a regular grid. To achieve this, a grid has been established over the metropolitan area of Milan. The grid resolution has been determined to ensure that only one monitoring station is located within each grid cell, with the station's value assigned to that cell. Consequently, the minimum distance between all stations was measured to identify the optimal resolution, which has been set at 1 km. However, given that the radius of influence for the stations is typically large and the measurements from the stations do not exhibit significant variability, in cases where it is unavoidable to have more than one station within a grid cell, the mean value of the measurements will be considered appropriate for representing the grid cell value.

3. Open Data Cube Implementation

The ODC (`https://www.opendatacube.org`) is an open-source tool designed to simplify working with large-scale satellite imagery and Earth observation data. It provides a structured framework for accessing, processing, and analyzing geospatial datasets over time, making it easier to track environmental observation patterns. The ODC helps transform raw geospatial data into ready-to-use datasets. It works by organizing gridded geospatial data into a multi-dimensional data structure (data cube) where each axis typically represents time, space, and additional variables like the data type. Data can be imported into the ODC from NetCDF or GeoTIFF formats. The ODC provides built-in API for data query, analysis, and download which can be accessed using Python scripting.

The ODC system provides a single end-point with analysis-ready data. For this project, the ODC helped the end-user to skip the data cleaning and processing procedure (Figure 3). This integration ensures structured data storage, enabling seamless querying and retrieval in an array-based format.

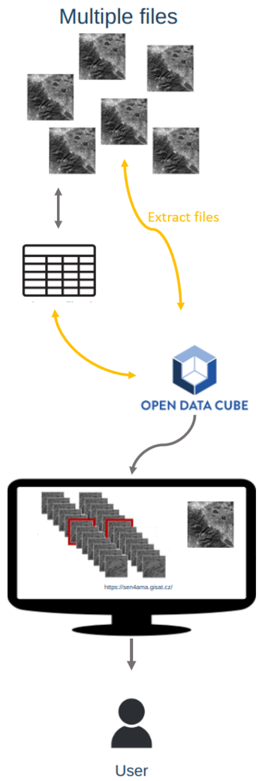

Figure 3. Schematic of the Open Data Cube data integration workflow.

The ODC integration began with the installation of the ODC in the local server. This consisted of setting up the ODC PostgreSQL databases, installing the related Python libraries and setting up the virtual environments from which the ODC will provide the end-point. Secondly, the ingestion process began with defining a new data Product in ODC. This Product acts as a logical container that groups datasets under a common structure, facilitating metadata assignment and data indexing. Although it does not store data itself, it plays a crucial role in query execution by serving as a reference layer for dataset management.

Next, a custom Python processing pipeline was developed to automate data preparation. This pipeline converts each CSV file into a NetCDF, format compatible with the Python Xarray library (`https://docs.xarray.dev`) and generates a corresponding YAML metadata file. The metadata file captures key attributes such as coordinate reference system, resolution, format, and content type. Once processed, both files are systematically registered in the ODC PostgreSQL database and linked to their respective Product. The same procedure was done with GeoTIFF files, which can be natively run inside the ODC.

Figure 4. Example of a Sentinel-5P ODC Product access and visualization through Python scripting.

Given the large volume of files used in this study, automation was essential to streamline the ingestion workflow. The ODC's API then facilitated spatial and temporal queries, allowing efficient data extraction using Product names as identifiers. The queried datasets returned as Xarray objects, enabled further statistical exploration of geospatial time series, as demonstrated in the subsequent analysis.

The implementation of the ODC will furnish the research community with a rigorous database sourced from various origins for air quality monitoring, while also offering user-friendly access to the community. This initiative will enable the integration of additional subject areas, allowing them to leverage the database for their specific applications. The establishment of the ODC will facilitate the development of a rigorously validated resource that can be continuously updated with new acquisitions, thereby enhancing its utility across diverse contexts. The code for ODC data preparation is available on GitHub (`https://github.com/gisgeolab/geoair_odc`).

4. Conclusion and Outlook

The developed ODC instance establishes a comprehensive air quality data management system by integrating data from various sources specific to the Metropolitan City of Milan. A significant advantage of the ODC system, in contrast to traditional methods of data downloading and processing, is its capacity to enable end-users (e.g., air quality monitoring agencies and researchers) to implement automated pipelines for data consumption within a unified platform. This functionality effectively reduces the time required for repetitive operations associated with data harmonization. Each data source utilized in this study has demonstrated its ability to capture a substantial signal relevant to the analysis of air quality phenomena. Although these sources may contain uncertainties or deficiencies in interpreting certain aspects, their integration has proven valuable, as they collectively encompass different dimensions of the air quality concept. The final outcome has been instrumental in enhancing awareness of the importance of air quality monitoring and can be utilized across various sectors. Furthermore, the developed methodology is highly adaptable and can be implemented in other regions, requiring only updates tailored to the specific area of interest.

The future direction of this research will concentrate on refining the source code and developing comprehensive documentation for the implemented ODC instance, with a particular emphasis on enhancing procedures for the retrieval and integration of multi-sensor data, including real-time capabilities. Given the limited number of ground sensors, a cost-effective network of sensors will be established to improve the diversity of coverage provided by ground measurements. These measurements are crucial for validating other data sources and enhancing the overall accuracy of the system. These advancements aim to facilitate the integration of the proposed framework into future air quality monitoring systems, thereby contributing to more effective urban air quality management and supporting the development of improved public health protection services.

5. ACKNOWLEDGEMENTS

This research was partially funded by the Italian Ministry of Education through the project of national interest (PRIN) "GeoIntelligence for improved air quality monitoring and analysis (GeoAIr)" (project code: 202258ACSL-PE10).

References

Bodah, B. W., Neckel, A., Maculan, L. S., Milanes, C. B., Korcelski, C., Ramírez, O., Mendez-Espinosa, J. F., Bodah, E. T., Oliveira, M. L., 2022. Sentinel-5P TROPOMI satellite application for NO2 and CO studies aiming at environmental valuation. *Journal of Cleaner Production*, 357, 131960.

Casciaro, G., Cavaiola, M., Mazzino, A., 2022. Calibrating the CAMS European multi-model air quality forecasts for regional air pollution monitoring. *Atmospheric environment*, 287, 119259.

Cedeno Jimenez, J. R., Oxoli, D., Brovelli, M. A., 2021. Enabling Air Quality Monitoring with the Open Data Cube: Implementation for Sentinel-5P and Ground Sensor Observations. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 46, 31–36.

Chen, G., Zhou, Y., Voogt, J. A., Stokes, E. C., 2024. Remote sensing of diverse urban environments: From the single city to multiple cities. *Remote Sensing of Environment*, 305, 114108.

Clelland, A. A., Marshall, G. J., Baxter, R., 2024. Evaluating the performance of key ERA-Interim, ERA5 and ERA5-Land climate variables across Siberia. *International Journal of Climatology*.

Copernicus Atmosphere Monitoring Service, 2025. CAMS regional: European air quality analysis and forecast data documentation. Available at: `https://ads.atmosphere.copernicus.eu`. Accessed on 10-02-2025.

Copernicus Climate Change Service, 2025. ERA5-Land hourly data from 1950 to present. `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land`. Accessed on 02-02-2025.

Dalla Torre, D., Di Marco, N., Menapace, A., Avesani, D., Righetti, M., Majone, B., 2024. Suitability of ERA5-Land reanalysis dataset for hydrological modelling in the Alpine region. *Journal of Hydrology: Regional Studies*, 52, 101718.

Douros, J., Eskes, H., van Geffen, J., Boersma, K. F., Compernolle, S., Pinardi, G., Blechschmidt, A.-M., Peuch, V.-H., Colette, A., Veefkind, P., 2023. Comparing Sentinel-5P TROPOMI NO 2 column observations with the CAMS regional air quality ensemble. *Geoscientific Model Development*, 16(2), 509–534.

European Environment Agency, 2023. Harm to human health from air pollution in Europe: Burden of disease 2023.

Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S., Francés, F., 2023. Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). *Atmospheric Research*, 284, 106606.

Lahoz, W., 2003. Atmospheric modelling. *Data Assimilation for the Earth System*, Springer, 149–166.

Lauritzen, P. H., Jablonowski, C., Taylor, M. A., Nair, R. D., 2011. *Numerical techniques for global atmospheric models*. 80, Springer Science & Business Media.

Maranzano, P., 2022. Air quality in Lombardy, Italy: an overview of the environmental monitoring system of ARPA Lombardia. *Earth*, 3(1), 172–203.

Meroni, G., Valerio, A., Vezzoli, M., Croci, E., Carruba, M. O., 2021. The relationship between air pollution and diabetes: A study on the municipalities of the Metropolitan City of Milan. *Diabetes Research and Clinical Practice*, 174, 108748.

Oxoli, D., Gianquintieri, L., Borghi, F., Fanti, G., Spinazzè, A., 2024. Non-Conventional Data for Farming-Related Air Pollution: Contributions to Modelling and Risk Assessment in the Lombardy Region, Italy. *Environments*, 11(11).

Parker, W. S., 2016. Reanalyses and observations: What's the difference? *Bulletin of the American Meteorological Society*, 97(9), 1565–1572.

Prunet, P., Lezeaux, O., Camy-Peyret, C., Thevenon, H., 2020. Analysis of the NO2 tropospheric product from S5P TROPOMI for monitoring pollution at city scale. *City and Environment Interactions*, 8, 100051.

Raffaelli, K., Deserti, M., Stortini, M., Amorati, R., Vasconi, M., Giovannini, G., 2020. Improving air quality in the Po Valley, Italy: some results by the LIFE-IP-PREPAIR project. *Atmosphere*, 11(4), 429.

Rafkin, S. C., Haberle, R. M., Michaels, T. I., 2001. The Mars regional atmospheric modeling system: Model description and selected simulations. *Icarus*, 151(2), 228–256.

Schneider, P., Hamer, P. D., Kylling, A., Shetty, S., Stebel, K., 2021. Spatiotemporal patterns in data availability of the sentinel-5p no2 product over urban areas in Norway. *Remote Sensing*, 13(11), 2095.

Sharma, S. B., Jain, S., Khirwadkar, P., Kulkarni, S., 2013. The effects of air pollution on the environment and human health. *Indian Journal of Research in Pharmacy and Biotechnology*, 1(3), 391–396.

Slivinski, L. C., 2018. Historical Reanalysis: What, How, and Why? *Journal of Advances in Modeling Earth Systems*, 10(8), 1736-1739.

Trini Castelli, S., Tinarelli, G., Reisin, T., 2017. Comparison of atmospheric modelling systems simulating the flow, turbulence and dispersion at the microscale within obstacles. *Environmental Fluid Mechanics*, 17, 879–901.

Virta, H., Ialongo, I., Szelag, M., Eskes, H., 2023. Estimating surface-level nitrogen dioxide concentrations from Sentinel-5P/TROPOMI observations in Finland. *Atmospheric Environment*, 312, 119989.

World Health Organization, 2021. *WHO Global Air Quality Guidelines*. World Health Organization, Geneva, Switzerland.

Zhang, Y., 2019. Dynamic effect analysis of meteorological conditions on air pollution: A case study from Beijing. *Science of the Total Environment*, 684, 178–185.

Zheng, Z., Yang, Z., Wu, Z., Marinello, F., 2019. Spatial variation of NO2 and its impact factors in China: An application of sentinel-5P products. *Remote Sensing*, 11(16), 1939.