ACCSCIENCE PUBLISHING

Asian Journal of Water, Environment and Pollution. 2026;23(2):025400308.
doi: 10.36922/AJWEP025400308

REVIEW ARTICLE

Advances in remote sensing and machine learning techniques for air quality monitoring

Alexander Uzhinskiy*

Meshcheryakov Laboratory of Information Technologies, Joint Institute for Nuclear Research, Dubna, Moscow Region, Russia
***Corresponding author:** Alexander Uzhinskiy (auzhinskiy@jinr.ru)

Received: October 2, 2025; Revised: November 17, 2025; Accepted: November 20, 2025; Published online: December 17, 2025

Abstract: Various techniques are used to assess air quality. Data for basic parameters, including particulate matter and major air pollutants, can be easily obtained from local meteorological stations, whereas obtaining more detailed information, such as heavy metal concentrations, requires laboratory analysis of collected samples. The structural limitations of regulatory monitoring systems limit their ability to provide continuous coverage across both the spatial and temporal domains. Satellite imagery provides vital information about atmospheric conditions and surface data. Each year, new missions with advanced sensors further enhance remote sensing capabilities. Sentinel-5 precursor, mounted with a tropospheric monitoring instrument, currently provides ready-to-use air quality data, including measurements of gases and aerosols, whereas Sentinel-4 and Sentinel-5, part of the Copernicus program, will extend these capabilities. Public satellite missions, such as Landsat and Sentinel, and instruments such as the moderate resolution imaging spectroradiometer, are widely used and provide high-resolution data with frequent updates. Integrating *in situ* measurements with satellite data and machine learning (ML) techniques enhances the accuracy and comprehensiveness of air quality monitoring. Modeling serves an important role in bridging data gaps, producing detailed assessments of particular areas, and supporting the partial automation of environmental control systems. This review assesses satellite-based programs, data processing tools, and project realization methods that enable efficient air quality estimation. The review demonstrates how ML-based remote sensing technology is effective for monitoring air quality, discusses commercial satellite missions, presents firsthand experience, and outlines future directions for advancing air quality monitoring technologies.

Keywords: Air quality monitoring; Remote sensing; Machine learning; Satellite missions; Environmental control; Modeling; Prediction

1. Introduction

Environmental monitoring has played a crucial role in human life since ancient times. The ancient civilizations of Egypt, China, Greece, and Rome observed environmental changes in water resources, air quality, deforestation, and other key parameters.[1-4] Numerous societies understood that environmental conditions directly affected public health. The Renaissance and Industrial Revolution enabled precise observation of the environment, leading to improved spatial and temporal data distributions. The 20th century introduced various international programs that began addressing environmental issues at scale. The Clean Air Act[5] of 1963 was the first federal legislation specifically targeted at air pollution control in the United States. It established a federal program within the U.S. Public Health Service and authorized research into techniques for monitoring and mitigating air pollution. The Geneva Convention on Long-Range Transboundary Air Pollution[6] became a significant milestone when it was adopted in 1979 through the United Nations Economic Commission for Europe (UNECE) framework and entered into force in 1983. The agreement allowed governments to work together on health and environmental protection from air pollution, affecting numerous countries.

Environmental monitoring programs now operate at multiple scales, including international and national levels, as well as regional, city, and enterprise levels, to provide comprehensive oversight of environmental challenges. This has become possible due to advances in digital technologies. The term Industry 4.0 represents industrial development transformation, whereas Enterprise 4.0, Healthcare 4.0, and Smart Environment Monitoring[7] describe analogous progress in their specific fields. Recent technological developments have brought substantial improvements to environmental monitoring and control systems. Internet of Things devices and sensors enable continuous data collection and transmission, whereas Big Data technologies enable the large-scale storage and processing of environmental data. Remote sensing technologies increasingly utilize both satellite and unmanned aerial vehicle platforms for environmental assessment. While the joint use of these systems is well established for soil[8] and water[9] quality monitoring, their combined use for air-pollution studies remains an open question. Digitalization and artificial intelligence (AI) enable efficient data analysis through strong tools for environmental modeling and prediction, which enhance decision-making and sustainability efforts.

This review focuses on aspects of air quality monitoring. Since air pollution is one of the leading environmental risk factors for premature death worldwide, this research area remains highly relevant. The Global Burden of Disease Study 2021 estimates that air pollution causes eight million deaths worldwide each year.[10] The methodology behind these estimates remains open to discussion, yet it is evident that poor air quality causes severe harm to human health. The standard method of air quality assessment depends on meteorological stations that monitor fundamental atmospheric data. In recent years, networks of low-cost, home-based air quality sensors have emerged. As part of larger networks, sensors create a comprehensive and precise spatial map of air pollution levels.[11] Ground-based stations with advanced capabilities, as well as mobile platforms mounted on vehicles or aircraft, can obtain detailed data on chemical composition and particle sizes. The spatial coverage of monitoring data remains limited even though these advances have been made, and modeling techniques are used to fill spatial gaps to produce a more comprehensive environmental picture.

To obtain specific, detailed information, researchers often rely on the laboratory-based analysis of collected samples. Various specialized techniques are used depending on the target pollutants. Moss biomonitoring, together with neutron activation analysis, serves as a common method for detecting airborne heavy metal contamination.[12] Mosses function as effective bioindicators because they absorb nutrients from the air rather than soil, reducing soil–metal interference, and their natural absorption properties make them suitable for detecting airborne pollutants. The effectiveness of laboratory-based methods is evident; however, they suffer from spatial and temporal limitations, including infrequent sampling, which occurs only once per year or less. Modeling functions enable the connection of unobserved data points to improve spatial and temporal detail and generate accurate continuous representations of environmental conditions.

Historically, a variety of modeling approaches have been employed to reconstruct the spatial and temporal distribution of air pollutants. The foundation of atmospheric pollutant dispersion and fate simulation relies on deterministic models that use equations for atmospheric transport and chemical transformations. The Weather Research and Forecasting model, coupled with chemistry, the Community Multiscale Air Quality model, and their variants, are widely used in regulatory applications and scenario-based forecasting, especially when coupled with high-resolution meteorological inputs and emission inventories.[13-15] The execution of these methods needs substantial computational power. They are complemented by statistical and data-driven models that have gained traction for their ability to estimate pollutant concentrations in areas lacking direct observations. Research methods, such as Kriging, Land Use Regression, and Bayesian hierarchical modeling, enable scientists to predict spatial patterns while using land use, traffic, and meteorological covariates.[16-18] These techniques offer flexibility and can perform well even with limited training data. The assessment of air quality now benefits from hybrid methods that combine deterministic modeling with statistical inference and machine learning (ML) techniques. These models enhance accuracy, reduce computational burden, and improve forecasting capacity by integrating deterministic physical structures with ML adaptive learning capabilities.[19,20]

Satellite remote sensing of the atmosphere began in the 1970s, when instruments, such as the Backscatter

Ultraviolet and Solar Backscatter Ultraviolet sensors on the Nimbus satellites, and later the Total Ozone Mapping Spectrometer, provided the first global measurements of ozone (O_3) concentration.[21] The launch of the Upper Atmosphere Research Satellite in 1991 and the Global Ozone Monitoring Experiment on European Remote Sensing Satellite 2 in 1995 extended these observations to include tropospheric trace gases, such as nitrogen dioxide (NO_2).[22] Building on these early developments, remote sensing data have become a valuable source of information for environmental modeling since the early 2000s. Modern multispectral and hyperspectral missions deliver data across multiple spectral bands with suitable spatial resolution and revisit frequency, enabling large-scale analyses of soil, water, vegetation, and atmospheric quality. Initially, most applications focused on land and water studies, but improvements in sensor technology, retrieval algorithms, and computational capacity have expanded their use to air-quality assessment. These applications primarily target key atmospheric pollutants—such as NO_2, sulfur dioxide (SO_2), O_3, carbon monoxide (CO), and methane (CH_4)—and retrieve aerosol optical depth (AOD), which is commonly used as an indirect indicator of particulate matter ($PM_{2.5}$).

The wider adoption of remote sensing in environmental studies remained restricted for years due to its demanding computational requirements and the need for experts who could handle geospatial software and satellite data analysis. The introduction of Google Earth Engine (GEE) platforms created a turning point by making large datasets, computational resources, and advanced algorithms accessible to researchers of all backgrounds, leading to a significant increase in studies using geographic information systems. Remote sensing and ML techniques are now standard tools for numerous air quality monitoring applications. These technologies enable researchers to predict pollutant concentrations and detect contamination sources, creating continuous spatial maps of air quality indicators.

This review selects recent research and interesting studies in the field instead of following the traditional scoping review format. The objective is to offer comprehensive information to those who wish to establish a solid base in this subject. The review provides basic information about relevant satellite programs, practical aspects of remote sensing data processing and modeling workflows, common challenges, and open research questions. The research involved is rooted in the development of a data management system for the UNECE International Cooperative Programme on Vegetation, as well as in the implementation of multiple projects that combined *in situ* air contamination data, remote sensing, and ML to predict at local, regional, and global scales. Recent research has applied statistical and Siamese neural network models to predict air quality in Central Russia.[23]

2. Key research developments in remote sensing and AI for air quality monitoring

The application of remote sensing and ML techniques for air quality monitoring has become an increasingly popular area of research. This emerging field has been extensively reviewed from various aspects. Holloway *et al.*[24] reviewed the expanding use of satellite data for air quality monitoring and its applications in public health and regulatory decision-making. Their 2021 review systematically summarizes the principal satellite products used to estimate pollutants, such as $PM_{2.5}$ and NO_2, details the integration of satellite observations with ground-based and modeled datasets, and examines a wide range of applications, including epidemiological studies, health impact assessments, wildfire monitoring, and policy support. Holloway *et al.*[25] extended this review by concentrating on the direct use of satellite-derived information in air quality assessment and management frameworks. The latter review provides an extensive evaluation of satellite product capabilities and limitations for regulatory use, while offering operational guidance on implementing remote sensing data in air quality planning, compliance evaluation, and long-term environmental policy development. Tang *et al.*[26] analyzed the present state of ML applications in air quality modeling with a critical evaluation. The review covers essential aspects, including feature engineering, data imbalance, model validation strategies, and interpretability. It also identifies methodological gaps and explores how these issues influence the reliability and conclusions of environmental models. A complementary perspective was provided by Binetti *et al.*,[27] who examined the deployment of supervised, unsupervised, and reinforcement learning across geoscientific domains. The review demonstrates the importance of these methods for environmental pattern recognition, prediction, and decision support and shows their transformative effect on both the scale and sophistication of environmental analyses, including air quality assessment. Garbagna *et al.*[28] presented a thorough systematic review of AI-based air pollution modeling techniques, including deep learning methods, hybrid models, and physics-informed ML approaches.

The review organizes methodologies according to input data types, model architectures, and target pollutants, and addresses challenges in model interpretability, data scarcity, and generalizability. It provides vital details on current developments and future opportunities for AI deployment in environmental pollution research. Stratoulias *et al.*[29] studied the recent advancements in satellite remote sensing for air pollution surveillance, emphasizing its critical role in supporting the United Nations Sustainable Development Goals. The research examines essential satellite missions, their data products, and methods for tracking pollutants, such as $PM_{2.5}$, NO_2, and SO_2. The paper demonstrates how satellite data are merged with ground observations and models and shows their practical applications in environmental policy and public health, especially in low- and middle-income countries.

In addition to these reviews, numerous high-quality research papers are published annually. A few notable examples are highlighted below. Shetty *et al.*[30] proposed a robust ML framework capable of estimating daily surface NO_2 concentrations for the European continent at 1 km spatial resolution. By integrating tropospheric monitoring instrument data from the Sentinel-5P satellite with auxiliary datasets, including the visible-infrared imaging radiometer suite (VIIRS) night-time lights, moderate resolution imaging spectroradiometer (MODIS)-derived vegetation indices, and meteorological variables, they achieved high predictive accuracy and leveraged Shapley additive explanations for model interpretability. The study underscores the utility of coupling remote sensing data with explainable AI methods to generate precise, high-resolution air quality maps. Rowley and Karakuş[31] introduced AQNet, a state-of-the-art multimodal AI architecture designed for air pollutant prediction. AQNet systematically fused raw multispectral satellite imagery from Sentinel-2 and satellite-derived NO_2 layers from Sentinel-5P with tabular environmental metadata, including population density, altitude, and station-specific parameters. The approach employed separate convolutional pathways to process image data before feature concatenation, enabling improved regression accuracy for predicting NO_2, O_3, and PM_{10} concentrations, even in areas with sparse ground measurements. Balamurugan *et al.*[32] explored the development and validation of gradient-boosted tree (GBT) and multilayer perceptron (MLP) models to estimate daily near-surface NO_2 and O_3 concentrations across Germany. By integrating diverse input features, including satellite data, meteorology, emission proxies, and vegetation indices, they demonstrated that GBT models outperformed both simpler MLP and conventional chemical transport models for NO_2 estimation, with the added benefit of reduced computational demand. The study further assessed the model's transferability to other geographical contexts. A study by Liang *et al.*[33] presents an advanced geospatial AI model that combines a calibrated low-cost sensor network, satellite-derived AOD, and light detection and ranging-based urban form data to predict daily $PM_{2.5}$ at 30 m resolution within a U.S. city. The attention-enhanced convolutional neural network architecture addresses urban-scale variability using multiscale residual kriging and permutation-based feature importance. The study highlights how integrating diverse data streams can enhance urban air pollution mapping and reveal spatial inequalities in exposure. Muthukumar *et al.*[34] proposed a deep-learning framework combining graph convolutional networks and convolutional long short-term memory (LSTM) to predict hourly $PM_{2.5}$ concentrations across Los Angeles. Using multisource inputs, including meteorological data, wildfire heat signatures, and remote-sensing imagery from MODIS and Sentinel-5P, the model achieves high spatiotemporal accuracy without relying on physical or chemical mechanisms. The study highlights the importance of integrating diverse environmental datasets for precise, real-time air quality forecasting. Chen *et al.*[35] applied ML algorithms to estimate urban air pollution based on remote sensing inputs. Their results show that models integrating meteorological, land-use, and multispectral data effectively support sustainable development strategies by providing actionable insights into air quality. Yang *et al.*[36] developed a nationwide $PM_{2.5}$ estimation model for China by fusing multiple satellite data sources, including MODIS, VIIRS, and Modern-Era Retrospective Analysis for Research and Applications, Version 2, with ML techniques. The resulting maps offer improved spatial coverage and accuracy for use in large-scale environmental monitoring. Malings *et al.*[37] proposed a data-fusion framework for real-time air quality estimation and forecasting that explicitly incorporated uncertainty quantification. By combining satellite observations, ground-based sensor data, and ML models, they improved the spatial and temporal resolution of $PM_{2.5}$ predictions while providing confidence intervals essential for policy-making and public health advisories. Xia *et al.*[38] presented a multimodal deep learning framework that fused remote sensing imagery with multi-station time-series data to predict air quality in Beijing and Tianjin. The method integrated spatial features from satellite data and temporal trends from ground monitoring networks using parallel convolutional and LSTM structures. The model demonstrated superior accuracy in estimating $PM_{2.5}$ concentrations compared to single-source baselines, emphasizing the benefits of multimodal data fusion for urban air pollution forecasting.

The selected review and research articles provide a comprehensive overview of current air quality monitoring innovations and practical methods, underscoring the need to combine remote sensing with ML and ground-based measurements to achieve robust, operationally relevant, and interpretable environmental assessments.

3. Satellite data

To develop effective environmental models, researchers require a combination of *in situ* and supplementary data that directly or indirectly correlate with environmental conditions. The integration process needs to establish clear causal relationships between features detected by remote sensing and ground-based environmental measurements. The most valuable sources of auxiliary data include satellite and aerial imagery, which allow researchers to extract quantitative features about land surface, atmospheric composition, vegetation, and pollution levels. The absorption and scattering of electromagnetic radiation across various spectral bands often provide information about the chemical composition or condition of the observed surface or atmosphere. Drone or crewed aircraft systems deliver precise spatial and temporal data through aerial imagery. The method has limitations, such as high cost, logistical challenges, and limited data availability. Open-access Earth observation programs enable environmental analysis using satellite imagery, providing global coverage with regular updates and multiple spectral bands. The described features enable satellites to function as strong tools for environmental monitoring and modeling. As of 2024, the National Aeronautics and Space Administration (NASA) reports that over 45,000 human-made objects orbit Earth, including more than 10,000 active satellites. This number has quadrupled since 2019.[39] Most satellites are designed for telecommunications and navigation purposes, yet multiple space missions can be used to monitor environmental conditions.

Notable examples include the Landsat program and the Copernicus Sentinel satellite series, mounted with MODIS and VIIRS. The missions provide various data products describing different atmospheric layers and land cover types; therefore, researchers can select data that matches their project needs. However, using optical satellite imagery for air quality monitoring presents inherent challenges. The main challenge arises from the requirement of clear-sky conditions since training datasets and predictions need to minimize cloud cover and atmospheric artifacts to achieve numerical accuracy. The process of creating cloud-free mosaics involves combining multiple images, but this method has limitations because satellites do not revisit their targets often enough. In contrast, synthetic aperture radar (SAR) sensors, such as those aboard Sentinel-1, actively emit microwave signals and can acquire data regardless of cloud cover or lighting conditions. Sensors used for SAR data acquisition are not well-suited for detecting gases, aerosols, and pollutants because SAR sensors capture geometric rather than spectral information.

The characteristics of the main Earth-observation missions and instruments used for environmental and atmospheric monitoring as of 2025 are summarized in Table 1. These include optical and thermal imagers, hyperspectral sensors, and atmospheric spectrometers from the Copernicus Programme, NASA, the National Oceanic and Atmospheric Administration, the European Organisation for the Exploitation of Meteorological Satellites, and the Korea Aerospace Research Institute.

Collectively, these missions provide complementary observations across spatial scales ranging from tens of meters to tens of kilometers and temporal resolutions from minutes to weeks. The near-daily revisit frequency of the satellite platforms carrying these instruments—Terra/Aqua (MODIS) and Suomi-NPP/NOAA-20 (VIIRS)—supports time-series analyses and facilitates cloud-free compositing despite their lower spatial resolution compared with Sentinel-2 and Landsat-9. In contrast, geostationary air-quality spectrometers—such as GEMS aboard GEO-KOMPSAT-2B and TEMPO hosted on Intelsat-40e—deliver hourly regional pollution observations. Sentinel-5 Precursor provides global and regional observations of key atmospheric pollutants, including NO_2, O_3, CO, CH_4, and SO_2. Numerous studies integrate Sentinel-5P data alone or fuse it with other satellite observations and ML algorithms, achieving notable gains in pollutant estimation accuracy. Nonetheless, spectral bands from different instruments often exhibit strong correlation. Researchers should also account for collinearity among closely related spectral bands, as this can reduce model robustness if not addressed through feature selection or dimensionality reduction techniques.

Table 1. Key satellite missions relevant to environmental and air-quality monitoring as of 2025

Mission/satellite	Operator/program	Spatial resolution (km)	Revisit time	Spectral range/bands	Primary pollutants or parameters observed
Landsat-8/9[40]	NASA/USGS	0.03	16 days	11 bands (VIS, NIR, SWIR)	Land cover, vegetation indices, TIR, surface temperature
Sentinel-2A/2B[41]	ESA Copernicus Programme	0.01–0.02	5 days	13 bands (VIS, NIR, SWIR)	Surface reflectance, vegetation health, AOD proxy
Terra/Aqua (mounted with MODIS)[42]	NASA	0.25–1.00	Daily	36 bands (VIS, TIR)	AOD, $PM_{2.5}$ proxy, fire hotspots, surface temperature
Suomi NPP/NOAA-20 (mounted with VIIRS)[43]	NOAA/NASA	0.375–0.750	Daily	22 bands (VIS, TIR)	AOD, night-time lights, thermal anomalies, cloud products
Sentinel-5 Precursor (mounted with TROPOMI)[44]	ESA Copernicus Programme	3.5×5.5	1–2 days	8 bands (UV, VIS, NIR, SWIR)	NO_2, SO_2, O_3, CO, CH_4, HCHO (global coverage)
Sentinel-4[45]	ESA Copernicus/ EUMETSAT (MTG)	~8	Hourly (Europe)	3 bands (UV, VIS, NIR)	NO_2, SO_2, O_3, formaldehyde (FCHO), aerosols
Sentinel-5[46]	ESA Copernicus Programme (Planned 2025)	7	Daily	7 bands (UV, VIS, SWIR)	NO_2, SO_2, CO, CH_4, O_3
GEO-KOMPSAT-2B (mounted with GEMS)[47]	KARI/NASA collaboration	7	Hourly (Asia)	Hyperspectral (UV, VIS; continuous spectral range)	NO_2, SO_2, O_3, HCHO, AOD (Asia region)
Intelsat-40e (mounted with TEMPO)[48]	NASA	~8	Hourly (North America)	Hyperspectral (UV, VIS; continuous spectrum, no discrete band count)	NO_2, SO_2, O_3, HCHO
Aura (mounted with OMI)[49]	NASA/KNMI	13×24	Daily	Hyperspectral (UV, VIS)	NO_2, SO_2, O_3, HCHO, AOD
MetOp C (mounted with GOME-2)[50]	EUMETSAT/DLR	40×80	Daily	Hyperspectral (UV-VIS)	NO_2, SO_2, O_3, BrO, HCHO

Abbreviations: AOD: Aerosol optical depth; DLR: German aerospace center; ESA: European space agency; EUMETSAT: European organisation for the exploitation of meteorological satellites; GEMS: Geostationary environment monitoring spectrometer; GEO-KOMPSAT: Geostationary Korea multipurpose satellite; GOME: Global ozone monitoring experiment; KARI: Korea aerospace research institute; MODIS: Moderate resolution imaging spectroradiometer; NASA: National aeronautics and space administration; NIR: Near-infrared; NOAA: National oceanic and atmospheric administration; NPP: National polar-orbiting partnership; OMI: Ozone monitoring instrument; $PM_{2.5}$: Particulate matter below 2.5 µm; SWIR: Short-wave infrared; TEMPO: Tropospheric emissions: Monitoring of pollution; TIR: Thermal infrared; TROPOMI: Tropospheric monitoring instrument; UV: Ultraviolet; VIIRS: Visible infrared imaging radiometer suite; VIS: Visible light.

While public programs offer excellent datasets for scientific research, commercial satellite operators provide extremely high-resolution imagery and rapid revisit cycles, albeit at high cost. The leading providers include:

- Planet Labs[51]: Operating large fleets (Dove, SkySat, Pelican, Hyperspectral), Planet Labs offers daily global coverage at resolutions up to 30 cm, with up to 12 revisits per day and 4–8 spectral bands. The new hyperspectral line, Tanager, provides ~30 m resolution and over 200 spectral bands for advanced environmental and climate applications.[52] Planet Labs' imagery supports agriculture, forestry, environmental monitoring, disaster response, and others.
- Maxar Technologies[53]: As a long-established company, Maxar deploys WorldView and GeoEye satellites with very high-resolution imaging (up to 30 cm) and up to 15 revisits daily across 29 spectral channels. Their data are favored for urban planning, defense, and infrastructure monitoring.
- Airbus Defence and Space[54]: Delivers data from its optical and radar Earth observation missions, including Pléiades, Pléiades Neo, Vision-1, SAR satellites, and Defense Meteorological Satellites. Airbus Pléiades Neo achieves up to 30 cm with daily intraday revisits and 5–7 bands.

If open-access satellite programs incorporate even a fraction of the capabilities currently offered by commercial providers—such as sub-meter spatial resolution or high temporal revisit frequencies—the effectiveness of data-driven environmental modeling would increase substantially. These advancements would support the long-term goal of scalable, near-real-time environmental monitoring. At the same time, ongoing developments in space-based Earth observation are already expanding global monitoring capacity. Upcoming missions, such as Sentinel-4, Sentinel-5, TANGO (The Netherlands/European Space Agency [ESA]),[55] Copernicus Carbon Dioxide Monitoring (ESA),[56] and Multi-Angle Imager for Aerosols,[57] are designed to deliver improved spatial resolution, enhanced spectral sensitivity, and more frequent observations for atmospheric pollutant analysis. Figure 1 summarizes NASA's current and planned missions under the Earth Science Division, illustrating how these efforts contribute to an integrated global system for operational air-quality assessment and long-term climate monitoring.

4. Satellite data processing

In the past, environmental monitoring through remote sensing methods needed both advanced technical

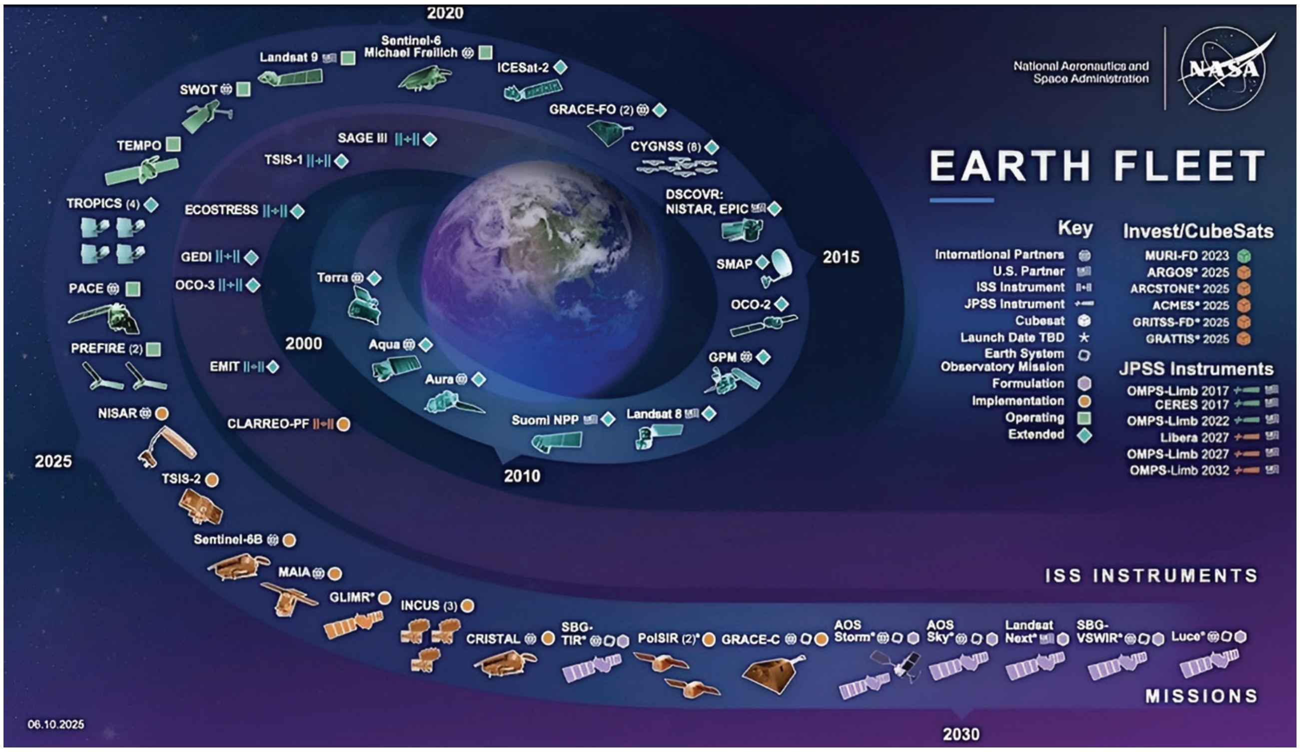

Figure 1. Overview of National Aeronautics and Space Administration, Earth Science Division satellite missions and data streams. Reprinted from Ref[58].

knowledge and powerful computing systems. Researchers had to manually select and download large volumes of imagery from public repositories. A single scene (tile) could easily exceed 1 GB of data, and constructing a continuous observation for a large region often involved numerous files. Pre-processing steps required specialized geospatial analysis platforms, such as Earth Resource Data Analysis System, Imagine, Environment for Visualizing Images, or Arc Geographic Information System, for atmospheric correction, geometric rectification, cloud masking, and band math operations. Only upon completing these labor-intensive tasks could the researcher proceed to analytics, visualization, feature extraction, and modeling. The process of moving from raw satellite data to operational models or applications could take weeks or months, limiting the field to experienced practitioners and relatively well-funded organizations.

GEE[59] became publicly available in 2013, establishing a new milestone. GEE combines a petabyte-scale catalog of satellite imagery with planetary-scale analysis tools in a cloud platform. The system has reduced entry requirements because researchers who know JavaScript or Python and basic remote sensing principles can now perform advanced geospatial analysis, satellite scene processing, index computation, and statistic extraction using Google's computational infrastructure. The democratization of earth observation analytics has led to an explosion of research output and innovation. Comparable cloud-based services include Sentinel Hub[60] and the Microsoft Planetary Computer (MPC).[61] All these platforms continue to evolve rapidly. The GEE and MPC platforms support open-source libraries and reproducible workflows across broad community ecosystems, whereas Sentinel Hub is particularly notable for its efficient real-time access to Sentinel-1, Sentinel-2, and Sentinel-5P imagery with customized visualization and application programming interface integration. These platforms enable real-time environmental monitoring, operational modeling, and scalable service development. For example, the GEE-based application, AlgaeMAp[62] demonstrates its ability to track harmful algal blooms in Latin American inland waters. The system generates outputs at 30 m spatial resolution and 5-day temporal resolution using the normalized difference chlorophyll index processing of Sentinel-2 imagery. The system enables real-time water quality assessments at particular locations through its functionality, which does not require local computing resources.

The platform provides technical capabilities to support various applications for atmospheric monitoring, emission mapping, and pollution trend analysis; however, there is a lack of dedicated online applications for air quality monitoring. It is likely that the field will develop increasingly advanced, practical applications in the near future.

5. Project realization aspects

Implementing environmental monitoring systems using remote sensing and ML requires systematic attention to multiple methodological aspects. The process should begin with clearly defined objectives that specify the ML task type, spatial and temporal coverage, and the main purpose, such as forecasting, exposure assessment, or informing environmental policy.

Different ML approaches are suitable for different types of environmental analysis. Regression tasks aim to predict continuous outputs, such as pollutant concentrations (e.g., $PM_{2.5}$, CO_2, and NO_2). In many cases, traditional models, such as gradient boosting, random forest regression, or support vector regression, provide competitive results, but an increasing number of studies are exploring deep learning approaches, particularly when good features and sufficient data are available.[63-65] Classification tasks involve labeling inputs into discrete classes (e.g., pollution levels: low, moderate, and high). Classification is generally more robust in environmental monitoring when ground-truth data are limited, as it enables the application of class-balancing techniques and offers access to a broad range of evaluation metrics (e.g., accuracy, recall, F1 score, and confusion matrices). In a large portion of research, statistical methods demonstrate strong performance; however, recent studies also show that advanced neural and ensemble approaches have strong potential, particularly in scenarios with limited training data.[66-70] Segmentation and clustering are essential for spatially explicit analyses, such as pollution source mapping, land use segmentation, or environmental region grouping. Statistical methods, such as K-Means and G-Means, along with neural network models, such as U-Net and DeepLabv3, are commonly used in Earth observation research.[71-73] Selecting the right task type and methodology should be based on a clear understanding of the problem and a review of the latest tools and frameworks in the field.

High-quality training data are critical. The quality of ground-truth measurements needs to be both representative, complete, and evenly distributed across space and time. Special care is warranted for rare yet high-impact pollution events, which are frequently underrepresented, potentially leading to an underestimation of model risk. The verification of outlier or extreme samples is essential; laboratory errors, local transient sources, or instrument artifacts must be rigorously cross-checked. Data from an area initially presumed to be clean could reveal unexpectedly high concentrations, which may later be traced to contaminant sources near the sampling site, a finding uncovered only after re-evaluating the location.

Satellite data play a crucial role in both model training and validation. Features extracted from satellite imagery must have a plausible cause for the target variable. For example, the direct relationship between *in situ* $PM_{2.5}$ and PM_{10} concentration and satellite-derived AOD has been corroborated in numerous studies.[74-76] Other predictors, such as night-time light intensity (e.g., from VIIRS or Defense Meteorological Satellite Program-Operational Linescan System), can serve as an indirect indicator for human activity and anthropogenic pollution sources.[77,78] Researchers should have a full understanding of the pipeline for extracting satellite-derived characteristics. Special attention must be given to several issues, such as cloud cover, orbital gaps, and retrieval errors, as these can introduce significant uncertainty into satellite-based observations if not properly pre-filtered or quality-screened. The size of the analyzed area also matters. For water or soil monitoring, pixel-level analysis is often the best choice. When investigating atmospheric characteristics, a broader area comprising multiple pixels can be analyzed. In such cases, spatially aware techniques, such as convolutional neural networks, may be employed,[79,80] or aggregation functions can be applied to derive combined characteristics. The median function is commonly used, although other operations, such as maximum or sum, may also be appropriate depending on the application.

Beyond satellite imagery, environmental ML models can benefit from integrating elevation data, meteorological variables (e.g., wind, humidity, and precipitation), land cover, traffic density, and similar contextual factors. Platforms such as GEE and MPC provide access to many of these datasets, simplifying integration. When incorporating static variables (e.g., land use), the indirect nature of their relationship with target variables must be acknowledged—these act more as contextual or moderating features than direct indicators. In the feature engineering process, it is always useful to assess the impact of each feature on the model's performance to better understand the obtained results.

Environmental data often exhibit class imbalance; rare pollution spikes or hazardous events may be an order of magnitude less prevalent than typical observations. To address this issue, resampling techniques or tailored loss functions can be employed. Among data-balancing methods, reducing the majority class is the safest but often impractical due to limited sample sizes. In such cases, synthetic oversampling techniques, such as the synthetic minority over-sampling technique and the adaptive synthetic sampling approach for imbalanced learning, can be adopted.[81]

Validation strategy is another core concern; random splits often lead to overly optimistic estimates due to spatial or temporal leakage. Instead, validation should use spatial and temporal cross-validation strategies to ensure realistic performance estimates. Especially with neural models, two models with similar metrics can exhibit substantial differences in prediction quality, potentially leading to hotspot drift or phantom risks. In such cases, the second level of model verification can be performed by comparing predicted hotspots with known or independently observed hotspots. Experts should choose reference locations to select hotspots that fall outside measurement areas, while maintaining predictable contamination ranges. Industrial areas, along with busy streets, often show elevated pollution levels; however, national parks and forest reserves should exhibit minimal contamination. After defining the list of expected hotspots and clean areas, the second step of model verification can be conducted. Only models that have demonstrated adequate performance in this verification step are used in subsequent stages of the research. The validation process serves to eliminate statistical models that are valid yet do not align with ecological reality.

Modeling requires strict adherence to documentation standards and version control mechanisms for code, data, parameter settings, and pre-processing workflows. The approach ensures scientific transparency and accelerates collaborative improvement by enabling independent validation. The promotion of reproducible research benefits from the use of public repositories, such as GitHub and Zenodo, and open notebooks, including Jupyter and RMarkdown.

To improve clarity and summarize the methodological workflow described in this section, Figure 2 provides a structured overview of the key stages involved in implementing a remote-sensing-based environmental monitoring system. The diagram outlines the sequence from defining objectives and collecting *in situ* measurements to satellite data pre-processing, feature engineering, model training, and final application deployment.

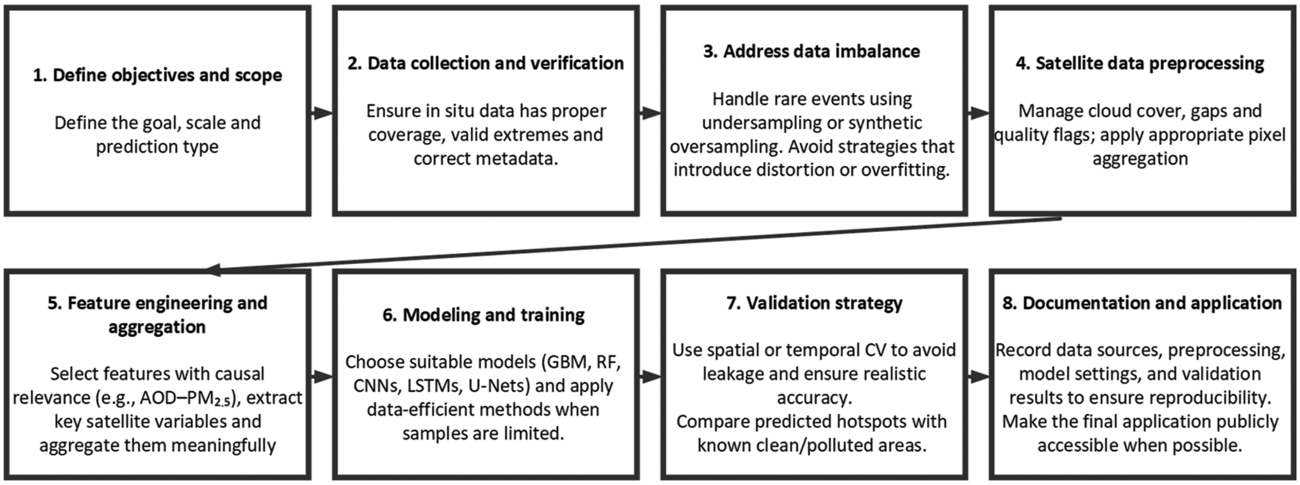

Figure 2. Workflow for implementing a data-driven environmental monitoring and control system using remote sensing and machine learning

Abbreviations: AOD: Aerosol optical depth; CNNs: Convolutional neural networks; CV: Cross-validation; GBM: Gradient boosting machine; LSTMs: Long short-term memories; PM: Particulate matter; RF: Random forest.

6. Conclusion

Air quality monitoring remains a critical area of environmental research. While global air pollution continues to pose substantial health and environmental risks, many regions are now demonstrating measurable progress through effective policy implementation and technological innovation. For instance, recent studies in Western Europe report nitrogen oxide emission reductions of up to 10% per year in urban areas, reflecting the positive impact of stricter air-quality regulations and cleaner transport technologies.[82] Continuous observation and assessment remain essential for sustaining these improvements and extending them to regions with less developed monitoring frameworks.

We are living in a time of significant technological progress with a growing number of tools available to assess the state of the environment. The combination of advanced sensor stations and effective estimation techniques enables precise *in situ* measurement of different air pollutants. However, challenges related to spatial and temporal coverage persist, particularly in regions with limited monitoring infrastructure. The combination of remote sensing technology with ML provides an effective solution to this problem, improving the resolution, scope, and responsiveness of air quality assessments. Looking forward, the next decade will likely see increased automation and near-real-time integration of diverse monitoring platforms, further narrowing data gaps and enhancing early warning systems. The progress will be driven by continued advances in satellite technology, data science, and AI. Raising awareness and understanding of these emerging capabilities is essential for researchers, practitioners, and decision-makers working in environmental monitoring.

Acknowledgments

The author gratefully acknowledges the support of the Meshcheryakov Laboratory of Information Technologies and the Frank Laboratory of Neutron Physics at the Joint Institute for Nuclear Research.

Funding

None.

Conflict of interest

The author declares no competing interests

Author contributions

This is a single-authored article.

Availability of data

Not applicable.

References

1. Jia S. Evolution of water governance in China. *J Water Resour Plan Manage*. 2021;147:04021050. doi: 10.1061/(ASCE)WR.1943-5452.0001420

2. Bernard S, McConnell J, Di Rita F, *et al.* An environmental and climate history of the roman expansion in Italy. *J Interdiscip Hist*. 2023;54:1-41. doi: 10.1162/jinh_a_01971
3. Willems H, Dahms JM, editors. *The Nile: Natural and Cultural Landscape in Egypt*. Bielefeld, Germany: Transcript Verlag; 2017.
4. Dong G, Wang L, Zhang DD, *et al.* Climate-driven desertification and its implications for the ancient Silk Road trade. *Clim Past.* 2021;17:1395-1407. doi: 10.5194/cp-17-1395-2021
5. United States Environmental Protection Agency (EPA). *Summary of the Clean Air Act*; 1963. Available from: https://www.epa.gov/laws-regulations/summary-clean-air-act [Last accessed on 2025 Nov 26].
6. United Nations Economic Commission for Europe (UNECE). *Convention on Long-Range Transboundary Air Pollution*. Available from: https://unece.org/environment-policy/air/about [Last accessed on 2025 Nov 26].
7. Ullo SL, Sinha GR. Advances in smart environment monitoring systems using IoT and sensors. *Sensors*. 2020;20:3113. doi: 10.3390/s20113113
8. Jia J, Chen C, Liu Q, Ding B, *et al*. Soil salinity monitoring model based on the synergistic construction of ground-UAV-satellite data. *Soil Use Manage*. 2024;40:e12980. doi: 10.1111/sum.12980
9. Cillero Castro C, Domínguez Gómez JA, Delgado Martín J, *et al*. An UAV and satellite multispectral data approach to monitor water quality in small reservoirs. *Remote Sens*. 2020;12:1514. doi: 10.3390/rs12091514
10. Institute for Health Metrics and Evaluation (IHME). *Global Burden of Disease Study 2021 (GBD 2021) Results*. Seattle, United States: IHME; 2024. Available from: https://www.healthdata.org/sites/default/files/2024-05/gbd_2021_booklet_final_2024.05.16.pdf [Last accessed on 2025 Nov 26].
11. Tariq H, Tariq S, Crescini D, Mnaouer AB. State-of-the-art low-cost air quality sensors, assemblies, calibration and evaluation for respiration-associated diseases: A systematic review. *Atmosphere*. 2024;15:471. doi: 10.3390/atmos15040471
12. Kosior G, Frontasyeva M, Ziembik Z, *et al.* The moss biomonitoring method and neutron activation analysis in assessing pollution by trace elements in selected polish national parks. *Arch Environ Contam Toxicol*. 2020;79:310-320. doi: 10.1007/s00244-020-00755-6
13. Grell GA, Peckham SE, Schmitz R, *et al*. Fully coupled “online” chemistry within the WRF model. *Atmos Environ*. 2005;39:6957-6975. doi: 10.1016/j.atmosenv.2005.04.027
14. Byun D, Schere KL. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl Mech Rev*. 2006;59:51-77. doi: 10.1115/1.2128636
15. Li W, Tang B, Campbell PC, *et al.* Updates and evaluation of NOAA’s online-coupled air quality model version 7 (AQMv7) within the Unified Forecast System. *Geosci Model Dev*. 2025;18:1635-1660. doi: 10.5194/gmd-18-1635-2025
16. Hoek G, Beelen R, de Hoogh K, *et al*. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Environ*. 2008;42:7561-7578. doi: 10.1016/j.atmosenv.2008.05.057
17. Zhang Y, Wang X, Wang J, Li Y, Chen L. Advances in air quality modeling and forecasting. *Environ Sci Ecotechnol*. 2020;2:100023. doi: 10.1016/j.glt.2020.11.001
18. Gianquintieri L, Oxoli D, Caiani EG, Brovelli MA. State-of-art in modelling particulate matter (PM) concentration: A scoping review of aims and methods. *Environ Dev Sustain*. 2024;27:5889–25911. doi: 10.1007/s10668-024-04781-5
19. Smith J, Wang L, Gupta R. A systematic survey of air quality prediction based on deep learning. *Alexandria Eng J*. 2024;93:128-141. doi: 10.1016/j.aej.2024.03.031
20. Li J, Li Y, Wang Z, *et al.* Next generation air quality models: Dynamical mesh, new insights into mechanism, datasets and applications. *Curr Pollut Rep.* 2025;11:25. doi: 10.1007/s40726-025-00355-9
21. Labow GJ, McPeters RD, Bhartia PK, Kramarova N. A comparison of 40 years of SBUV measurements of column ozone with data from the dobson/brewer network. *J Geophys Res Atmos*. 2013;118:7370-7378. doi: 10.1002/jgrd.50503
22. Irie H, Sudo K, Akimoto H, *et al*. Evaluation of long-term tropospheric NO_2 data obtained by GOME over East Asia in 1996–2002. *Geophys Res Lett*. 2005;32:L11810. doi: 10.1029/2005GL022770
23. Uzhinskiy A, Vergel K. Central russia heavy metal contamination model based on satellite imagery and machine learning. *Comput Opt*. 2023;47:137-151. doi: 10.18287/2412-6179-CO-1149
24. Holloway T, Miller D, Anenberg S, *et al.* Satellite monitoring for air quality and health. *Annu Rev Biomed Data Sci*. 2021;4:417-447. doi: 10.1146/annurev-biodatasci-110920-093120
25. Holloway T, Bratburd JR, Fiore AM, Kerr GH, Mao J. Satellite data to support air quality assessment and management. *J Air Waste Manag Assoc*. 2025;75(6):429-463. doi: 10.1080/10962247.2025.2484153
26. Tang D, Zhan Y, Yang F. A review of machine learning for modeling air quality: Overlooked but important issues. *Atmos Res*. 2024;300:107261.

doi: 10.1016/j.atmosres.2024.107261
27. Binetti MS, Massarelli C, Uricchio VF. Machine learning in geosciences: A review of complex environmental monitoring applications. *Mach Learn Knowl Extr*. 2024;6:1263-1280. doi: 10.3390/make6020059
28. Garbagna L, Saheer LB, Maktab Dar Oghaz M. AI-driven approaches for air pollution modelling: A comprehensive systematic review. *Environ Pollut*. 2025;373:125937. doi: 10.1016/j.envpol.2025.125937
29. Stratoulias D, Nuthammachot N, Dejchanchaiwong R, Tekasakul P, Carmichael GR. Recent developments in satellite remote sensing for air pollution surveillance in support of sustainable development goals. *Remote Sens*. 2024;16:2932. doi: 10.3390/rs16162932
30. Shetty S, Schneider P, Stebel K, Hamer PD, Kylling A, Berntsen TK. Estimating surface NO_2 concentrations over europe using sentinel-5P TROPOMI observations and machine learning. *Remote Sens Environ*. 2024;312:114321. doi: 10.1016/j.rse.2024.114321
31. Rowley A, Karakuş O. Predicting air quality via multimodal AI and satellite imagery. *Remote Sens Environ*. 2023;293:113609. doi: 10.1016/j.rse.2023.113609
32. Balamurugan V, Chen J, Wenzel A, Keutsch FN. Spatiotemporal modeling of air pollutant concentrations in Germany using machine learning. *Atmos Chem Phys*. 2023;23:10267-10285. doi: 10.5194/acp-23-10285-2023
33. Liang L, Daniels J, Phillips R, South J, Hu L. Integrating low-cost sensor monitoring, satellite mapping, and geospatial artificial intelligence for intraurban air pollution predictions. *Environ Pollut*. 2023;331:121832. doi: 10.1016/j.envpol.2023.121832
34. Muthukumar P, Nagrecha K, Comer D, *et al.* PM2.5 air pollution prediction through deep learning using multisource meteorological, wildfire, and heat data. *Atmosphere*. 2022;13:822. doi: 10.3390/atmos13050822
35. Chen W, Zhang N, Bai X, Cao X. Research on the estimation of air pollution models with machine learning in urban sustainable development based on remote sensing. *Sustainability*. 2024;16:10949. doi: 10.3390/su162410949
36. Yang Y, Wang Z, Cao C, *et al*. Estimation of PM2.5 concentration across china based on multi-source remote sensing data and machine learning methods. *Remote Sens*. 2024;16:467. doi: 10.3390/rs16030467
37. Malings C, Emma Knowland K, Pavlovic N, *et al*. Air quality estimation and forecasting via data fusion with uncertainty quantification. *J Geophys Res Atmos*. 2024;1:e2024JH000183. doi: 10.1029/2024JH000183
38. Xia H, Chen X, Wang Z, Chen X, Dong F. A multi-modal deep-learning air quality prediction method based on multi-station time-series data and remote-sensing images: Case study of Beijing and Tianjin. *Entropy*. 2024;26:91. doi: 10.3390/e26010091
39. NASA. *NASA Make Way for Satellites*. Available from: https://spinoff.nasa.gov/make_way_for_satellites [Last accessed on 2025 Nov 26].
40. NASA/USGS. *Landsat 8 Overview*. Available from: https://landsat.gsfc.nasa.gov/landsat-8 [Last accessed on 2025 Nov 26].
41. European Space Agency (ESA). *Sentinel-2*. Available from: https://www.earthdata.nasa.gov/data/instruments/sentinel-2-msi [Last accessed on 2025 Nov 26].
42. NASA. *MODIS Overview*. Available from: https://modis.gsfc.nasa.gov/about [Last accessed on 2025 Nov 26].
43. NOAA/NASA. *VIIRS*. Available from: https://viirsland.gsfc.nasa.gov [Last accessed on 2025 Nov 26].
44. European Space Agency (ESA). *Sentinel-5P Mission Overview*. Available from: https://sentinels.copernicus.eu/copernicus/sentinel-5p [Last accessed on 2025 Nov 26].
45. European Space Agency (ESA). *Sentinel-4 Mission Overview*. Available from: https://www.esa.int/applications/observing_the_earth/copernicus/sentinel-4 [Last accessed on 2025 Nov 26].
46. European Space Agency (ESA). *Sentinel-5 Mission Overview*. Available from: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-5 [Last accessed on 2025 Nov 26].
47. National Institute of Environmental Research (NIER). *GEMS Overview*. Available from: https://nesc.nier.go.kr/en/html/cntnts/73/satellite/introduction.do [Last accessed on 2025 Nov 26].
48. Zoogman P, Liu X, Suleiman RM, *et al*. Tropospheric emissions: Monitoring of pollution (TEMPO). *J Quant Spectrosc Radiat Transf*. 2017;186:17-39. doi: 10.1016/j.jqsrt.2016.05.028
49. Levelt PF, van den Oord GHJ, Dobber MR, *et al*. The ozone monitoring instrument. *IEEE Trans Geosci Remote Sens*. 2006;44:1093-1101. doi: 10.1109/TGRS.2006.872333
50. German Aerospace Center (DLR). *GOME-2/MetOp Overview*. Available from: https://atmos.eoc.dlr.de/app/missions/gome2 [Last accessed on 2025 Nov 26].
51. Planet Labs. *Planet Labs Official Site*. Available from: https://www.planet.com [Last accessed on 2025 Nov 26].
52. Planet Labs. *Introducing Planet Tanager - Hyperspectral Satellite*. Available from: https://www.planet.com/pulse/planet-launches-first-tanager-1-hyperspectral-satellite-and-36-superdoves-with-spacex [Last accessed on 2025 Nov 26].
53. Maxar Technologies. *WorldView and GeoEye Satellite*

Overview. Available from: https://www.maxar.com/products/satellite-imagery [Last accessed on 2025 Nov 26].
54. Airbus Defence and Space. *Satellite Imagery Product Portal*. Available from: https://www.intelligence-airbusds.com [Last accessed on 2025 Nov 26].
55. TNO. *TANGO Satellite: Earth Observation for Greenhouse Gas Emissions Monitoring*; 2025. Available from: https://www.tno.nl/en/sustainable/earth-observation/tango-satellite [Last accessed on 2025 Nov 26].
56. European Space Agency (ESA). *Copernicus Carbon Dioxide Monitoring (CO_2M) Mission Overview*; 2025. Available from: https://www.eoportal.org/satellite-missions/co2m [Last accessed on 2025 Nov 26].
57. NASA Jet Propulsion Laboratory (JPL). *MAIA Mission Overview — Multi-Angle Imager for Aerosols*; 2024. Available from: https://maia.jpl.nasa.gov [Last accessed on 2025 Nov 26].
58. NASA Earth Science Division. *NASA Earth Fleet: Earth-Observing Satellite and Instrument Portfolio (Mission Timeline Poster)*. Available from: https://svs.gsfc.nasa.gov/30065 [Last accessed on 2025 Nov 26].
59. *Google Earth Engine Platform*. Available from: https://earthengine.google.com [Last accessed on 2025 Nov 26].
60. Sinergise. *Sentinel Hub Platform*. Available from: https://www.sentinel-hub.com [Last accessed on 2025 Nov 26].
61. Microsoft. *Planetary Computer Platform*. Available from: https://planetarycomputer.microsoft.com [Last accessed on 2025 Nov 26].
62. Lobo FdL, Nagel GW, Maciel DA, *et al*. AlgaeMAp: Algae bloom monitoring application for inland waters in Latin America. *Remote Sens*. 2021;13:2874. doi: 10.3390/rs13152874
63. Méndez M, Merayo MG, Núñez M. Machine learning algorithms to forecast air quality: A survey. *Artif Intell Rev*. 2023;56:10031-10066. doi: 10.1007/s10462-023-10424-4
64. Abuouelezz W, Ali N, Aung Z, *et al*. Exploring PM2.5 and PM10 ML forecasting models: A comparative study in the UAE. *Sci Rep.* 2025;15:9797. doi: 10.1038/s41598-025-94013-1
65. Ajala AA, Adeoye OL, Salami OM, *et al.* An examination of daily CO_2 emissions prediction through a comparative analysis of machine learning, deep learning, and statistical models. *Environ Sci Pollut Res*. 2025;32:2510-2535. doi: 10.1007/s11356-024-35764-8
66. Lilhore UK, Simaiya S, Singh RK, *et al.* Advanced air quality prediction using multimodal data and dynamic modelling. *Sci Rep.* 2025;15:27867. doi: 10.1038/s41598-025-11039-1
67. Fu Q, Guo H, Gu X*, et al*. High-resolution PM2.5 concentrations estimation based on stacked ensemble learning model using multi-source satellite TOA data. *Remote Sens*. 2023;15:5489. doi: 10.3390/rs15235489
68. Wei J, Wang Z, Li Z, *et al.* Global aerosol retrieval over land from landsat imagery integrating transformer and google earth engine. *Remote Sens Environ*. 2024;315:114404. doi: 10.1016/j.rse.2024.114404
69. Emeç M, Yurtsever M. A novel ensemble machine learning method for accurate air quality prediction. *Int J Environ Sci Technol*. 2025;22:459-476. doi: 10.1007/s13762-024-05671-z
70. Mampitiya L, Rathnayake N, Hoshino Y, Rathnayake U. Forecasting PM_{10} levels in Sri Lanka: A comparative analysis of machine learning models. *J Hazard Mater Adv*. 2024;13:100395. doi: 10.1016/j.hazadv.2023.100395
71. Park MH, Choi JH, Lee WJ. Toward sustainable clean ports: Smoke opacity measurement based on semantic segmentation and edge device. *J Environ Manage*. 2025;393:127139. doi: 10.1016/j.jenvman.2025.127139
72. Nurseitov D.B, Abdimanap G, Abdallah A, *et al.* A remote sensing satellite data for oil spill detection on land. *Eng Sci*. 2024;32:1348. doi: 10.30919/es1348
73. Gharahbagh AA, Hajihashemi V, Machado JJM, Tavares JMRS. Land cover classification model using multispectral satellite images based on a deep learning synergistic semantic segmentation network. *Sensors (Basel)*. 2025;25:1988. doi: 10.3390/s25071988
74. van Donkelaar A, Martin RV, Brauer M, *et al*. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ Health Perspect*. 2010;118(6):847-855. doi: 10.1289/ehp.0901623
75. Stirnberg R, Cermak J, Andersen H. An analysis of factors influencing the relationship between satellite-derived AOD and ground-level PM10. *Remote Sens*. 2018;10:1353. doi: 10.3390/rs10091353
76. Zhu H, Martin RV, van Donkelaar A, *et al*. Importance of aerosol composition and aerosol vertical profiles in global spatial variation in the relationship between PM2.5 and aerosol optical depth. *Atmos Chem Phys*. 2024;24:11565-11584. doi: 10.5194/acp-24-11565-2024
77. Li X, Zhou Y, Zhao M, *et al.* A harmonized global nighttime light dataset 1992–2018. *Sci Data*. 2020;7:168. doi: 10.1038/s41597-020-0510-y
78. Zheng Q, Seto KC, Zhou Y, You S, Weng Q. Nighttime light remote sensing for urban applications: Progress, challenges, and prospects. *ISPRS J Photogramm Remote Sens*. 2023;202:125-141. doi: 10.1016/j.isprsjprs.2023.05.028
79. Kayastha SG, Ghahremanloo M, Park J, Singh D,

Westenbarger D, Choi Y. A deep learning framework for satellite-derived surface $PM_{2.5}$ estimation: Enhancing spatial analysis in the United States. *Artif Intell Earth Syst*. 2024;3(4):e240028.
doi: 10.1175/AIES-D-24-0028.1

80. Zhang Z, Ao Z, Wu W, Wang Y, Xin Q. Developing a multi-scale convolutional neural network for spatiotemporal fusion to generate MODIS-like data using AVHRR and landsat images. *Remote Sens*. 2024;16:1086.
doi: 10.3390/rs16061086

81. Wongvorachan T, He S, Bulut O. A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. *Information*. 2023;14:54.
doi: 10.3390/info14010054

82. Verhoelst T, Compernolle S, Lambert JC, Vanpoucke C, Fierens F. Synergistic use of satellite and *in-situ* data for policy-relevant air quality information: A case study on Belgium. *Atmos Environ*. 2025;361:121447.
doi: 10.1016/j.atmosenv.2025.121447