Short-Term Forecasting of Ground-Level O_3 and NO_2: A Synthesis of Satellite and Reanalysis Data Applications

Executive Summary

The field of air quality forecasting is undergoing a profound transformation, driven by a confluence of high-resolution satellite remote sensing, comprehensive atmospheric reanalysis products, and the escalating power of advanced computational models. This report provides an exhaustive synthesis of the current state of research in short-term forecasting of ground-level ozone (O3) and nitrogen dioxide (NO2), two pollutants with significant impacts on public health and the environment. The analysis reveals a clear paradigm shift away from traditional, sparsely monitored systems toward integrated, data-driven frameworks capable of providing high-resolution, actionable intelligence.

A central catalyst for this evolution is the advancement in satellite observation capabilities. The transition from polar-orbiting instruments with single daily overpasses, such as the Ozone Monitoring Instrument (OMI), to high-resolution successors like the TROPOspheric Monitoring Instrument (TROPOMI) has enabled the resolution of pollution at the urban scale. More critically, the advent of geostationary platforms, exemplified by the Geostationary Environment Monitoring Spectrometer (GEMS), provides hourly observations that capture the diurnal cycles of short-lived pollutants, a capability that is revolutionizing short-term forecast modeling.

These rich data streams are contextualized by atmospheric reanalysis products, such as NASA's MERRA-2 and the Copernicus Atmosphere Monitoring Service (CAMS) reanalysis. These datasets provide the essential, physically consistent meteorological information—including wind fields, temperature, and planetary boundary layer height—that governs the transport and chemical transformation of pollutants. They serve as the indispensable connective tissue that allows models to bridge the critical gap between satellite-observed total column densities and the ground-level concentrations relevant to human exposure.

The modeling paradigms themselves have evolved in lockstep with data availability. While statistical methods like Land Use Regression (LUR) have been significantly enhanced by satellite data, they are increasingly being surpassed by machine learning (ML) and deep learning (DL) models. Algorithms such as Random Forest, XGBoost, and particularly time-aware architectures like Long Short-Term Memory (LSTM) networks, have demonstrated

superior performance in capturing the complex, non-linear relationships inherent in atmospheric systems. Case studies of novel DL systems like GeoNet and O3ResNet show they can outperform even state-of-the-art operational forecast ensembles.

Despite these advances, significant scientific and operational challenges persist. The primary scientific hurdle remains the accurate and reliable estimation of surface concentrations from satellite-retrieved column densities. This requires sophisticated modeling and the robust integration of meteorological data. Furthermore, issues of data integrity, including gaps from cloud cover and the harmonization of multi-resolution datasets, demand advanced preprocessing and data fusion techniques. Finally, a "valley of death" often exists between high-performing research models and their operational implementation, a gap that can only be bridged by rigorous, purpose-driven validation protocols that assess not just overall accuracy, but a model's fitness for specific applications, such as predicting exceedances of health-based air quality standards.

The future of air quality forecasting lies in the development of hybrid, AI-driven systems. These frameworks will synergistically fuse multi-source, near-real-time data streams from the next generation of geostationary satellites with high-resolution numerical weather predictions. By doing so, they will provide not just forecasts, but a comprehensive air quality intelligence system capable of supporting dynamic public health advisories, evaluating policy effectiveness, and identifying emission sources with unprecedented precision.

The Observational Backbone: Satellite Instruments and Reanalysis Datasets

The capacity to forecast ground-level air quality is fundamentally dependent on the quality, resolution, and frequency of observational data. In recent decades, two categories of data products have emerged as the cornerstones of this effort: direct observations of atmospheric composition from satellite-based remote sensing instruments and comprehensive, gridded atmospheric state variables from reanalysis datasets. The co-evolution of these data sources has directly enabled the development of increasingly sophisticated forecasting models.

Revolution in Remote Sensing: From Polar-Orbiting to Geostationary Satellites

The ability to monitor atmospheric pollutants from space has progressed from providing coarse, long-term global snapshots to delivering high-resolution, high-frequency regional data streams. This technological trajectory has been pivotal, as the type and complexity of feasible forecasting models are directly dictated by the characteristics of the available satellite data.

The Legacy and Evolution: OMI and TROPOMI for NO_2 and O_3 Monitoring

The modern era of satellite-based air quality monitoring was inaugurated by instruments like the Ozone Monitoring Instrument (OMI), which has been operational on NASA's Aura satellite since 2004.[1] OMI provided the first consistent, long-term global dataset for key pollutants like NO2 and O3, enabling foundational research into global pollution trends and the identification of large-scale emission hotspots. However, its utility for local, short-term forecasting was constrained by its relatively coarse spatial resolution of 13 km × 24 km at nadir.[2] This resolution was insufficient to resolve fine-scale pollution gradients within urban areas.
A significant leap forward occurred with the launch of the TROPOspheric Monitoring Instrument (TROPOMI) aboard the European Space Agency's Sentinel-5 Precursor (S5P) satellite in 2017.[2] TROPOMI offers an order-of-magnitude improvement in spatial resolution, with pixels as small as 5.5 km × 3.5 km, and a superior signal-to-noise ratio.[5] This enhanced capability allows for the detection of pollution sources at the scale of individual cities, power plants, and major transportation corridors, making it a cornerstone for contemporary air quality analysis and modeling.[8] Studies directly comparing the two instruments confirm that while their measurements are highly correlated (spatial correlation coefficient
R>0.93), TROPOMI's fine resolution provides a vastly superior representation of spatial variability and local pollution details, which are often missed by OMI due to its larger pixel size and issues like the "row anomaly" that degrades data quality.[2] The availability of TROPOMI data has directly enabled more powerful statistical models; for instance, incorporating TROPOMI
NO2 columns into a Land Use Regression (LUR) model over the United States improved the adjusted R2 from 0.54 to 0.72, a far greater impact than that of any land-use variable.[8]

The Temporal Frontier: The Impact of Geostationary Instruments like GEMS

While TROPOMI revolutionized the spatial dimension of air quality monitoring, both it and OMI share a fundamental limitation inherent to their polar-orbiting nature: they observe a given location on Earth only once per day, typically in the early afternoon (~13:30 local time).[1] This single daily snapshot is often insufficient for short-term forecasting, as it cannot capture the full diurnal evolution of pollutants. This is especially problematic for reactive species like NO2, which has a short atmospheric lifetime of only a few hours during the day.[1] Its concentration can vary significantly from the morning rush hour to the afternoon photochemical peak, a dynamic that once-daily measurements cannot resolve.
Geostationary satellites represent a paradigm shift in addressing this temporal limitation. By orbiting in sync with the Earth's rotation, they can stare continuously at a specific region, providing frequent measurements throughout the daylight hours. The Geostationary Environment Monitoring Spectrometer (GEMS), launched by South Korea in 2020, is the world's first UV-Vis spectrometer in a geostationary orbit dedicated to air quality.[1] GEMS

provides hourly daytime measurements of
NO2, O3, SO2, and other pollutants over a large portion of East Asia.[1]
This high-frequency data provides an unprecedented observational constraint for forecasting models. Instead of a single data point, models can now be trained on a time series of observations, allowing them to learn the dynamic processes of emission, chemical transformation, and transport that occur throughout the day.[1] This capability is a critical enabler for the next generation of accurate, data-driven short-term air quality forecasts, particularly those based on time-aware deep learning architectures that require sequential data for training.[1]

Instrument	Satellite Platform	Operational Period	Orbit Type	Revisit Time	Spatial Resolution (Nadir)	Key Pollutant Products
OMI	NASA Aura	2004–Present	Polar, Sun-synchronous	Daily	13 km × 24 km	NO2, O3, SO2, HCHO, Aerosols
TROPOMI	ESA Sentinel-5P	2017–Present	Polar, Sun-synchronous	Daily	~5.5 km × 3.5 km	NO2, O3, SO2, CO, CH4, HCHO, Aerosols
GEMS	GEO-KOMPSAT-2B	2020–Present	Geostationary	Hourly (daytime)	3.5 km × 7.5 km	NO2, O3, SO2, HCHO, Aerosols

Contextualizing the Atmosphere: The Role of Reanalysis Data

Satellite instruments provide crucial information on the abundance and distribution of pollutants, but this information exists in a vacuum without atmospheric context. The formation, transport, and dispersion of pollutants are governed by meteorological conditions. Reanalysis datasets provide this context by combining a vast array of historical observations with a modern weather model to produce a physically consistent, spatially complete, gridded dataset of the state of the atmosphere over several decades. These products are not merely sources of weather data; they are the essential connective tissue that allows for the physical and statistical integration of sparse ground-based measurements and column-integrated satellite observations.
A satellite measures the total amount of a pollutant in a vertical column of the atmosphere, but health impacts are determined by concentrations at the surface.[14] The link between these two quantities is the vertical mixing depth of the atmosphere, most often characterized by the Planetary Boundary Layer Height (PBLH).[13] Reanalysis models, by assimilating data from weather balloons, aircraft, and surface stations, provide the most reliable, spatially continuous

estimates of PBLH and other variables like wind speed, temperature, and humidity. They thus serve as the indispensable physical translator that enables forecasting models to convert what the satellite sees (a total column) into what is needed for public health applications (a surface concentration).

NASA's MERRA-2: Assimilating Aerosols and Meteorological Variables

The Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) is a global atmospheric reanalysis produced by NASA's Global Modeling and Assimilation Office (GMAO). It provides a continuous and consistent climate record from 1980 to the present at a horizontal resolution of approximately 50 km (0.5° × 0.625°).[15] MERRA-2 provides the full suite of meteorological variables required for air quality modeling, including temperature, pressure, humidity, and wind vectors at various atmospheric levels.[16] A key innovation of MERRA-2 is that it was the first major long-term reanalysis to directly assimilate space-based observations of aerosols.[15] This provides valuable data on the distribution of particulate matter and its chemical components (e.g., black carbon, dust, sea salt), which are often co-emitted with and interact chemically with gaseous pollutants like
NO2 and O3 precursors.[17]

The Copernicus Atmosphere Monitoring Service (CAMS): A European Perspective on Global Composition

The Copernicus Atmosphere Monitoring Service (CAMS), implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides a suite of products focused specifically on atmospheric composition.[22] The CAMS global reanalysis (currently designated EAC4) covers the period from 2003 to the present at a resolution of approximately 80 km.[24] Unlike purely meteorological reanalyses, CAMS assimilates a wide array of satellite retrievals of chemical species to produce comprehensive three-dimensional fields of pollutants including
O3, NO2, carbon monoxide (CO), and aerosols.[25] CAMS data serves multiple purposes: it can be used as a standalone forecast product, as a source of chemical boundary conditions for higher-resolution regional models, or as a benchmark for evaluating other models.[22] While CAMS provides a robust global and regional baseline, case studies have shown that its forecasts can exhibit local-scale biases. For example, one study found that CAMS underestimated wintertime particulate matter concentrations in a Mediterranean city, likely due to an underrepresentation of local emissions from residential wood burning.[29] This highlights a common theme: while global reanalysis and forecast products are invaluable, achieving high accuracy at the local level often requires post-processing or data fusion with local observations.

Methodological Paradigms in Air Quality Forecasting

The transformation of raw observational data from satellites and reanalysis systems into accurate short-term forecasts is accomplished through a diverse array of modeling techniques. The field has seen a clear and logical progression in methodology, moving from simpler statistical models designed for spatial interpolation to highly complex machine and deep learning architectures capable of capturing intricate, non-linear spatiotemporal dynamics. This evolution reflects a maturing understanding of the air quality forecasting problem, with the choice of model being dictated by the complexity of the available data and the specific nature of the forecasting task.

Statistical Foundations: From Regression to Geostatistical Inference

Early approaches to leveraging disparate data sources for air quality mapping and forecasting were rooted in statistical methods. These models form an important baseline and continue to be refined with the advent of higher-quality data.

- **Land Use Regression (LUR):** LUR models are a widely used statistical technique that predicts pollutant concentrations at a given location based on surrounding land-use and geographic variables, such as proximity to roads, population density, and industrial activity.[8] The integration of satellite data as a predictor variable has proven to be a major advancement for LUR. A study demonstrated that including TROPOMI-derived NO2 column densities in an LUR model for the United States dramatically improved its predictive power, increasing the adjusted R2 from 0.54 to 0.72. The satellite data proved to be a more influential predictor than any other variable, including road networks, underscoring the value of direct atmospheric observations.[8]
- **Bayesian and Time-Series Approaches:** Geostatistical methods provide another avenue for fusing data from multiple sources. The Bayesian Maximum Entropy (BME) framework, for instance, can integrate sparse ground-based monitor data with spatially continuous satellite retrievals to generate high-resolution (1×1 km) maps of pollutant concentrations, effectively filling the gaps between monitoring stations.[32] For forecasting temporal changes, time-series models like the Seasonal Autoregressive Integrated Moving Average (SARIMA) have been applied. These models excel at identifying and projecting historical patterns, such as weekly cycles and seasonal variations, present in pollutant data from ground stations or satellites.[34] While powerful for capturing temporal autocorrelation, traditional statistical models may struggle to represent the highly non-linear chemical and physical interactions that govern pollutant concentrations.[35]

The Machine Learning Ascendancy: Capturing Non-Linear Dynamics

The proliferation of large, multi-dimensional datasets from satellites and reanalysis has spurred a rapid shift towards machine learning (ML) and deep learning (DL) techniques. A comprehensive review of the literature reveals a significant surge in the application of these methods since 2021, with ML/DL now dominating the field.[37] Their primary advantage lies in their ability to learn complex, non-linear relationships directly from data without requiring explicit programming of physical and chemical laws.[38]

- **Ensemble Methods (Random Forest, XGBoost):** Ensemble tree-based models are among the most popular and effective ML approaches. Random Forest (RF) is frequently cited for its high accuracy and robustness, with some studies reporting prediction accuracies as high as 98.2%.[37] Extreme Gradient Boosting (XGBoost) has also emerged as a top-performing algorithm, particularly for forecasting O3 and NO2. In a comparative study for Beijing, XGBoost achieved the highest accuracy for O3 prediction, with an R2 of 0.767 when using meteorological and pollutant variables.[41] In another innovative application, a nested XGBoost model was developed to first predict the NO2 mixing height (NMH) from meteorological data, and then use that predicted NMH to estimate ground-level NO2 from GEMS satellite columns, achieving an exceptionally high cross-validation R2 of 0.93.[12]
- **Deep Learning Architectures (CNNs, LSTMs):** As the forecasting problem has become increasingly spatiotemporal with the advent of geostationary data, deep learning models have become more prevalent. These architectures are designed to automatically extract relevant features from complex data structures.
    - **Convolutional Neural Networks (CNNs)** are well-suited for processing gridded, image-like data. By applying convolutional filters, they can learn spatial patterns from satellite images or reanalysis fields. One study found that a simple CNN architecture, when applied directly to grids of satellite land-use data, significantly outperformed traditional LUR-style models (including RF and linear regression) for predicting daily ground-level NO2, achieving an R2 of 0.892.[31]
    - **Long Short-Term Memory (LSTM)** networks are a type of Recurrent Neural Network (RNN) specifically designed to learn dependencies in sequential data, making them ideal for time-series forecasting.[42] Studies consistently show that LSTMs and their variants, such as Bidirectional LSTMs (BiLSTMs), excel when temporal patterns are critical for prediction.[41]
- **Key Case Studies:** Two recent deep learning systems exemplify the state-of-the-art:
    - **GeoNet:** This neural network was designed to leverage the high-frequency observations from the GEMS geostationary satellite. By training on spatiotemporal series of satellite NO2 data, GeoNet can produce a 24-hour forecast for surface NO2 over eastern China. It demonstrated a strong performance, with an R2 of 0.68 against ground-based measurements, significantly surpassing traditional air quality models.[1]
    - **O3ResNet:** This deep learning system, based on a residual CNN, was developed

to produce a 4-day forecast for ground-level O3. Trained on a massive 22-year dataset of in-situ measurements and ERA5 reanalysis data for central Europe, O3ResNet was shown to outperform the state-of-the-art CAMS regional forecast model ensemble in terms of mean-square error and mean absolute error, marking a major milestone for DL-based operational forecasting.[45]

A notable tension exists within the field between the pursuit of predictive performance and the need for model interpretability. Deterministic and simple statistical models are transparent—their outputs can be traced to specific inputs or equations. In contrast, high-performing ML and DL models often function as "black boxes," making it difficult to understand the reasoning behind a specific forecast. This lack of transparency can be a significant barrier to their adoption in regulatory and policy contexts, where understanding the "why" is as crucial as the "what." This challenge has given rise to an emerging focus on eXplainable AI (XAI), which employs techniques to probe these complex models and elucidate feature importance, thereby building the trust necessary for operational deployment.[27]

Deterministic Approaches: The Physics and Chemistry of Chemical Transport Models (CTMs)

Distinct from data-driven statistical and ML models, Chemical Transport Models (CTMs) are deterministic, process-based simulations of the atmosphere. Models like the Community Multiscale Air Quality Model (CMAQ), GEOS-Chem, and WRF-Chem simulate the full lifecycle of pollutants—including emissions, advection and diffusion, complex chemical reactions, and deposition—based on the fundamental laws of physics and chemistry.[46] They provide a comprehensive, three-dimensional, and physically consistent representation of atmospheric composition.

However, CTMs have significant limitations. They are computationally intensive, which constrains their operational spatial resolution and the speed at which forecasts can be generated.[49] More importantly, their accuracy is fundamentally dependent on the quality of their input data, particularly emission inventories, which are often uncertain, incomplete, or outdated.[48] These input errors can lead to persistent systematic biases in CTM outputs when compared to real-world observations.[46] For example, CTMs often struggle to accurately simulate the rapid chemical evolution and the ratio of
NO to NO2 within fresh emission plumes, leading to discrepancies when comparing model outputs to satellite NO2 retrievals.[46]

The Power of Synergy: Hybrid Models and Data Fusion Frameworks

Recognizing that no single modeling paradigm is perfect, the leading edge of research is increasingly focused on developing hybrid frameworks that combine the strengths of different approaches to overcome their individual weaknesses.[50]

- **Data Fusion:** This strategy involves the statistical combination of data from multiple sources to produce a single, improved estimate. A powerful application of this is the fusion of satellite observations, CTM simulations, and ground-based monitor data.[50] In this approach, the CTM provides a spatially and temporally complete initial estimate of the pollutant field. This "first guess" is then corrected and refined using the high spatial coverage of satellite data and the high accuracy of sparse ground-based measurements. This method leverages the CTM's physical consistency while correcting its inherent biases with real-world observations.[50] The synergistic use of satellite and reanalysis data as inputs to an AI model is another potent form of data fusion.[27]
- **Model Hybridization:** This involves combining different types of models. For instance, a hybrid forecasting system might use a statistical model like SARIMA to capture the linear, seasonal components of a pollutant time series, and then use an LSTM network to model the remaining complex, non-linear patterns in the data.[35] Another sophisticated technique involves data decomposition. A method using wavelet decomposition first separates the air quality time series into its high-frequency (volatile) and low-frequency (trend) components. Each component is then modeled with the most appropriate tool—an LSTM for the complex high-frequency signal and a simpler ARMA model for the smoother low-frequency signal. The final forecast is reconstructed by summing the predictions for each component, resulting in significantly higher accuracy than either model could achieve alone.[53]

Study/Model Name	Pollutant	Model Type	Key Data Inputs	Performance (R2)	Performance (RMSE)	Region
TROPOMI LUR [8]	NO2	LUR	TROPOMI, Land Use	0.72 (Adj.)	-	United States
CNN [31]	NO2	CNN	Satellite Land Use	0.892 (Daily)	2.26 µg/m3	United States
GeoNet [1]	NO2	Neural Network	GEMS, ERA5, CAMS	0.68	12.31 µg/m3	Eastern China
Nested XGBoost [13]	NO2	XGBoost	GEMS, Meteorology	0.93 (CV)	-	China
O3ResNet [45]	O3	CNN (Residual)	In-situ, ERA5	Outperforms CAMS	Outperforms CAMS	Central Europe
XGBoost-LF PM [41]	O3	XGBoost	In-situ, Meteorology	0.873	8.17 µg/m3	Beijing, China
Data Fusion [50]	PM2.5	Hybrid (LME, Kriging)	AOD, CMAQ, Monitors	0.72	23.0 µg/m3	China

Critical Challenges in Operational Forecasting

Transitioning a forecasting model from a research environment to a reliable, operational system presents a distinct set of scientific and practical challenges. High performance on a curated historical dataset does not guarantee utility for real-world decision-making. Key hurdles include accurately relating satellite column measurements to surface air quality, managing data imperfections, and implementing validation protocols that are truly fit for an operational purpose.

The Vertical Challenge: Bridging the Gap from Column Density to Surface Concentration

The most significant conceptual challenge in using satellite data for health-relevant air quality applications is the fundamental difference between what a satellite measures and what is required for exposure assessment. Satellite instruments typically retrieve the total vertical column density (VCD), which is the integrated number of pollutant molecules in a column of air from the ground to the top of the atmosphere.[14] However, health impacts and regulatory standards are based on the concentration of pollutants in the air at the surface, where people live and breathe.[14]

The relationship between VCD and surface concentration is not fixed; it is highly dynamic and is primarily governed by the vertical structure of the atmosphere. The key variable is the pollutant's mixing height, which is closely related to the planetary boundary layer (PBL) height.[13] On a day with a deep, well-mixed boundary layer, pollutants are dispersed through a large volume of air, resulting in lower surface concentrations for a given VCD. Conversely, on a day with a shallow boundary layer (e.g., under a temperature inversion), pollutants are trapped near the ground, leading to high surface concentrations even with a moderate VCD. Therefore, accurately estimating ground-level concentrations requires robust information on the state of the PBL. This is where reanalysis data becomes critical. A study using GEMS data over China developed a nested XGBoost model that explicitly tackled this problem. The model's first stage predicted the NO2 mixing height (NMH) using meteorological variables from reanalysis. This predicted NMH was then fed as a crucial input into the second stage, which predicted surface NO2 from the satellite VCD. This two-step approach dramatically improved model performance, with the cross-validation R2 increasing from 0.73 to 0.93. A feature importance analysis revealed that the NMH was the second most important predictor of surface concentration, surpassed only by the satellite VCD itself.[13] This result powerfully illustrates that the "vertical challenge" can be overcome, but only through the synergistic integration of satellite observations with meteorological data from reanalysis or high-quality numerical weather prediction models. This dependency also implies that the ultimate accuracy of any satellite-based air quality forecast is fundamentally limited by the accuracy of the underlying weather forecast that provides these critical meteorological inputs.

The Data Integrity Challenge: Addressing Gaps, Clouds, and

Resolution Mismatches

Operational forecasting systems must be robust to the imperfections inherent in real-world data streams. Optical satellite instruments, which form the backbone of NO2 and O3 monitoring, are unable to see through clouds. This is a major source of data gaps, with studies suggesting that cloud cover can obscure anywhere from 40% to 80% of potential observations, leading to significant spatiotemporal discontinuities in the data record.[56]

Effective data preprocessing is therefore a critical first step in any forecasting pipeline. Simple methods for handling missing values, such as filling with a daily or monthly mean, are often inadequate for time-series applications as they distort the temporal structure of the data. More sophisticated techniques, such as linear interpolation or spatiotemporal kriging, are required to fill gaps in a more physically plausible manner.[57]

A further challenge lies in harmonizing the disparate spatial and temporal resolutions of the various input datasets. A typical forecasting model might need to ingest point-based data from ground monitors (continuous in time, sparse in space), gridded satellite data at a ~5 km resolution (spatially continuous but with a single daily or hourly snapshot), and coarser gridded reanalysis data at a ~50-80 km resolution (spatially and temporally continuous).[27] The process of resampling, averaging, and interpolating these different data types onto a common modeling grid is a non-trivial step that can introduce significant uncertainty into the final forecast.[57] The fundamental mismatch between a point measurement from a ground monitor and a grid-cell-averaged value from a satellite or model remains a persistent source of error and an active area of research.[14]

The Validation Imperative: Performance Metrics and Benchmarking

Validating the performance of a forecasting model is essential for establishing its credibility and fitness for purpose. While standard statistical metrics such as the coefficient of determination (R2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) are widely used and provide a valuable top-level assessment of model accuracy, they do not tell the whole story.[31]

A robust validation strategy must go further. It should employ techniques like k-fold cross-validation to ensure that the model is not simply "memorizing" the training data and can generalize to new, unseen data. For air quality, station-based (or leave-one-out) cross-validation is particularly important, as it tests a model's ability to make predictions at locations for which it has no direct training data, simulating a key use case of filling observational gaps.[40]

Even these methods can fall short of assessing true operational utility. A model can have a high overall R2 while still failing at the task that matters most for public health: predicting extreme pollution events. For regulatory and public health applications, the ability of a model to correctly forecast exceedances of air quality standards is paramount. This has led to the development of more purpose-driven validation frameworks, such as the one proposed by the Forum for Air quality Modeling in Europe (FAIRMODE).[61] This approach uses a suite of statistical and categorical metrics designed to specifically evaluate a model's capability to detect sudden changes in concentration, predict threshold exceedances, and accurately reproduce air quality indices.[61] The gap between the high accuracies often reported in academic research papers and the more modest performance of operational systems like CAMS when evaluated in real-world settings highlights this "valley of death" between research and operations. Bridging this gap requires a shift in focus from simply optimizing for metrics like RMSE to a more holistic, purpose-driven validation that proves a model is reliable, robust, and truly fit for its intended operational purpose.

Synthesis and Future Outlook

The short-term forecasting of ground-level O3 and NO2 has matured into a highly dynamic and data-intensive field. The convergence of advanced satellite platforms, comprehensive reanalysis systems, and powerful machine learning algorithms has established a new state-of-the-art and set a clear trajectory for future development. The path forward points toward increasingly integrated, AI-driven systems that leverage near-real-time data streams to deliver high-resolution air quality intelligence for a wide range of societal applications.

A Synthesized View of the Current State-of-the-Art

The most advanced and successful contemporary approaches to short-term air quality forecasting are fundamentally hybrid in nature. They synergistically combine the unique strengths of multiple data sources and modeling techniques to create a system that is more powerful than the sum of its parts. An archetypal state-of-the-art system leverages:

- **High-resolution satellite data** from polar-orbiting instruments like TROPOMI to capture fine-scale spatial gradients of pollutants.
- **High-frequency satellite data** from geostationary instruments like GEMS to observe the diurnal evolution of pollution.
- **Physically consistent meteorological context** from reanalysis products like MERRA-2 or CAMS to model the transport and chemical environment.
- **Indispensable ground-truth** from surface monitoring networks for model training, validation, and bias correction.
- **The predictive power of machine learning algorithms**, such as XGBoost or LSTM networks, to learn the complex, non-linear relationships between these disparate data streams and produce an accurate forecast.

The selection of a specific modeling architecture involves navigating a critical set of trade-offs. Simpler statistical models are computationally efficient and highly interpretable but may lack the accuracy needed for operational use. Deterministic CTMs offer unparalleled physical and chemical detail but are computationally prohibitive and highly sensitive to uncertainties in their input data, particularly emission inventories. ML and DL models consistently deliver the highest predictive accuracy but often at the cost of interpretability, creating a "black box" that can be challenging to trust for regulatory decisions. The optimal choice is therefore not universal but depends on the specific application, available computational resources, the need for interpretability, and the required forecast resolution and lead time.

The Next Generation of Air Quality Intelligence

The field is poised for another leap forward, enabled by a new generation of observational assets and modeling capabilities. The imminent launch of geostationary air quality monitoring satellites over North America (NASA's TEMPO) and Europe (ESA's Sentinel-4) will extend the high-frequency observational capacity pioneered by GEMS to the rest of the developed world.[63] These missions, along with new satellites specifically designed for health applications like NASA's MAIA, will generate an unprecedented volume and velocity of data, further solidifying the role of AI-driven models as the primary tool for analysis and forecasting.[55]

In parallel, global forecasting systems are pushing the boundaries of spatial resolution. Models like NASA's GEOS Composition Forecast (GEOS-CF) are on a clear path to increase their native resolution from ~25 km to ~12 km and beyond, with the long-term goal of matching the performance of regional models while providing seamless global coverage.[65]

The ultimate trajectory of the field is toward the development of fully integrated air quality intelligence systems, akin to a "digital twin" of the atmosphere. These will not be standalone models but dynamic frameworks that:

1. Continuously assimilate near-real-time data from a constellation of geostationary and polar-orbiting satellites.[1]
2. Are driven by high-resolution numerical weather prediction forecasts for the most accurate meteorological context.[65]
3. Employ a suite of AI and ML models for forecasting, data fusion, and real-time bias correction.[43]
4. Deliver a portfolio of tailored products beyond simple concentration maps, including targeted public health alerts for vulnerable populations, rapid-response analysis of extreme events like wildfires, robust tools for policy scenario evaluation, and data-driven methods for identifying and quantifying emission sources.[55]

To realize this future, the research and operational communities should prioritize several key areas. First, continued investment in open-access, analysis-ready data infrastructure is paramount to ensure that the vast datasets from new satellite missions are readily usable by the global scientific community. Second, there is an urgent need to develop and adopt standardized, "fit-for-purpose" validation protocols for ML-based forecast models to build trust and facilitate their transition into operational use. Finally, fostering deeper interdisciplinary collaboration between atmospheric scientists, data scientists, public health professionals, and policymakers is essential to ensure that the next generation of forecasting tools is not only technically advanced but also directly addresses the most pressing needs of society.

Works cited

1. Unleashing the potential of geostationary satellite observations in air quality forecasting through artificial intelligence techniques - ACP, accessed September 22, 2025, https://acp.copernicus.org/articles/25/759/2025/
2. Comparison and Validation of TROPOMI and OMI NO2 Observations over China - MDPI, accessed September 22, 2025, https://www.mdpi.com/2073-4433/11/6/636
3. NO2 Retrieval from the Environmental Trace Gases Monitoring Instrument (EMI): Preliminary Results and Intercomparison with OMI and TROPOMI - MDPI, accessed September 22, 2025, https://www.mdpi.com/2072-4292/11/24/3017
4. ARSET - High Resolution NO2 Monitoring From Space with TROPOMI | NASA Applied Sciences, accessed September 22, 2025, https://appliedsciences.nasa.gov/get-involved/training/english/arset-high-resolution-no2-monitoring-space-tropomi
5. TROPOMI Observing Our Future | TROPOMI Observing Our Future | TROPOMI: TROPOspheric Monitoring Instrument, accessed September 22, 2025, https://www.tropomi.eu/
6. A comparison of the impact of TROPOMI and OMI ... - AMT, accessed September 22, 2025, https://amt.copernicus.org/articles/15/1703/2022/amt-15-1703-2022.pdf
7. tropomi - Tropospheric Monitoring Instrument - NASA Earthdata, accessed September 22, 2025, https://www.earthdata.nasa.gov/data/instruments/tropomi
8. TROPOMI Satellite Data Reshape NO2 Air Pollution Land-Use ..., accessed September 22, 2025, https://pubs.acs.org/doi/10.1021/acsestair.4c00153
9. The three ways of comparing the CAMS NO2 forecasts with TROPOMI. In the... - ResearchGate, accessed September 22, 2025, https://www.researchgate.net/figure/The-three-ways-of-comparing-the-CAMS-NO2-forecasts-with-TROPOMI-In-the-top-row-the_fig1_367411262
10. (PDF) Comparison and Validation of TROPOMI and OMI NO2 Observations over China, accessed September 22, 2025, https://www.researchgate.net/publication/342223227_Comparison_and_Validation_of_TROPOMI_and_OMI_NO2_Observations_over_China
11. HAQAST Sentinel-5P TROPOMI Nitrogen Dioxide (NO2) CONUS Monthly Level 3 0.01 x 0.01 Degree Gridded Data V2.4 (HAQ_TROPOMI_NO2_CONUS_M_L3) at GES DISC - CMR Search, accessed September 22, 2025, https://cmr.earthdata.nasa.gov/search/concepts/C2839237275-GES_DISC.html
12. Estimation of ground-level NO2 and its spatiotemporal variations in China using GEMS measurements and a nested machine learning model - ACP, accessed September 22, 2025, https://acp.copernicus.org/articles/24/9645/
13. Estimation of ground-level NO2 and its spatiotemporal variations in China using GEMS measurements and a nested machine learning - ACP, accessed September 22, 2025,

https://acp.copernicus.org/articles/24/9645/2024/acp-24-9645-2024.pdf
14. Satellite Monitoring for Air Quality and Health - NASA Technical ..., accessed September 22, 2025, https://ntrs.nasa.gov/api/citations/20230000905/downloads/Holloway_review_satellite%20monitoring%20for%20air%20quality%20and%20health.pdf
15. Modern-Era Retrospective analysis for Research and ... - GMAO, accessed September 22, 2025, https://gmao.gsfc.nasa.gov/gmao-products/merra-2/
16. The NASA MERRA-2 Reanalysis Products: Data and Tools Used for Aerosol and Air Quality Studies, accessed September 22, 2025, https://ntrs.nasa.gov/api/citations/20230008307/downloads/Poster_ACAM2023_MERRA2_AQ_xpan_poster.pdf
17. NASA's MERRA2 reanalysis - Climate Data Guide, accessed September 22, 2025, https://climatedataguide.ucar.edu/climate-data/nasas-merra2-reanalysis
18. What (and How) MERRA-2 Reanalysis Data are Used in Applied Sciences, accessed September 22, 2025, https://ntrs.nasa.gov/api/citations/20210026850/downloads/AGU_iPoster_shen_202112.pdf
19. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) - PMC - PubMed Central, accessed September 22, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC6999672/
20. The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part II: Evaluation ..., accessed September 22, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7477811/
21. Spatial and Temporal Distribution of PM 2.5 Pollution over Northeastern Mexico: Application of MERRA-2 Reanalysis Datasets - MDPI, accessed September 22, 2025, https://www.mdpi.com/2072-4292/12/14/2286
22. Air quality - Copernicus Atmosphere Monitoring Service, accessed September 22, 2025, https://atmosphere.copernicus.eu/air-quality
23. Characteristics of the atmospheric composition reanalysis products CAMS and MERRA-2., accessed September 22, 2025, https://www.researchgate.net/figure/Characteristics-of-the-atmospheric-composition-reanalysis-products-CAMS-and-MERRA-2_tbl1_391906120
24. CAMS Reanalysis | ECMWF, accessed September 22, 2025, https://www.ecmwf.int/en/research/climate-reanalysis/cams-reanalysis
25. The CAMS reanalysis of atmospheric composition - Semantic Scholar, accessed September 22, 2025, https://pdfs.semanticscholar.org/88e8/acb827819168fedef3d34af0e205a3a877b1.pdf
26. CAMS global reanalysis - ECMWF, accessed September 22, 2025, https://www.ecmwf.int/en/forecasts/dataset/cams-global-reanalysis
27. Estimation of Daily Ground Level Air Pollution in Italian Municipalities with Machine Learning Models Using Sentinel-5P and ERA5 Data - MDPI, accessed September 22, 2025, https://www.mdpi.com/2072-4292/16/7/1206
28. Use cases - Copernicus Atmosphere Monitoring Service, accessed September 22, 2025, https://atmosphere.copernicus.eu/use-cases
29. Forecasting Particulate Pollution in an Urban Area: From ... - MDPI, accessed

September 22, 2025, https://www.mdpi.com/2073-4433/12/7/881

30. Technical note: Accurate, reliable, and high-resolution air quality predictions by improving the Copernicus Atmosphere Monitoring Service using a novel statistical post-processing method - ACP, accessed September 22, 2025, https://acp.copernicus.org/articles/24/1673/2024/
31. National ground-level NO2 predictions via satellite imagery driven convolutional neural networks - Frontiers, accessed September 22, 2025, https://www.frontiersin.org/journals/environmental-science/articles/10.3389/fenvs.2023.1285471/full
32. Spatio-Temporal Prediction of Ground-Level Ozone Concentration Based on Bayesian Maximum Entropy by Combining Monitoring and Satellite Data - MDPI, accessed September 22, 2025, https://www.mdpi.com/2073-4433/13/10/1568
33. Intercomparison of global ground-level ozone datasets for health- relevant metrics - EGUsphere, accessed September 22, 2025, https://egusphere.copernicus.org/preprints/2025/egusphere-2024-3723/egusphere-2024-3723.pdf
34. Current Status and Future Forecast of Short-lived Climate-Forced Ozone in Tehran, Iran, derived from Ground-Based and Satellite Observations - PubMed Central, accessed September 22, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9930078/
35. Advanced Hybrid Models for Air Pollution Forecasting: Combining SARIMA and BiLSTM Architectures - MDPI, accessed September 22, 2025, https://www.mdpi.com/2079-9292/14/3/549
36. Air Quality Forecast by Statistical Methods: Application to ... - Frontiers, accessed September 22, 2025, https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2022.826517/full
37. Application of artificial intelligence in air pollution ... - PoliTO, accessed September 22, 2025, https://iris.polito.it/retrieve/dae9ccdd-aa3b-4fe7-9d81-41f8426777d7/%282024%29%20paper%20-%20review%20air%20pollution.pdf
38. Data-Driven Framework for Understanding and Predicting Air Quality in Urban Areas, accessed September 22, 2025, https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2022.822573/full
39. A Review of Machine Learning Models for Predicting Air Quality in Urban Areas - ASPG, accessed September 22, 2025, https://www.americaspg.com/article/pdf/3492
40. Using remotely sensed data for air pollution assessment - arXiv, accessed September 22, 2025, https://arxiv.org/html/2402.06653v1
41. Comparison of machine learning methods for predicting ... - Frontiers, accessed September 22, 2025, https://www.frontiersin.org/journals/environmental-science/articles/10.3389/fenvs.2025.1561794/full
42. A Novel Evolutionary Deep Learning Approach for PM 2.5 Prediction Using

Remote Sensing and Spatial–Temporal Data: A Case Study of Tehran - MDPI, accessed September 22, 2025, https://www.mdpi.com/2220-9964/14/2/42
43. Forecasting of Air Quality with Machine Learning, accessed September 22, 2025, https://2025.iaia.org/final-papers/1261_Makala_Forecasting_of_air_Quality.pdf
44. A hybrid model for daily air quality index prediction and its performance in the face of impact effect of COVID-19 lockdown - PubMed Central, accessed September 22, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10264166/
45. O3ResNet: A Deep Learning–Based Forecast System to Predict Local Ground-Level Daily Maximum 8-Hour Average Ozone in Rural and Suburban Environments in - AMS Journals, accessed September 22, 2025, https://journals.ametsoc.org/view/journals/aies/2/3/AIES-D-22-0085.1.xml
46. Potential Errors in CMAQ NO:NO2 Ratios and Upper Tropospheric NO2 Impacting the Interpretation of TROPOMI Retrievals, accessed September 22, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC12172011/
47. (PDF) Air quality modelling, simulation, and computational methods ..., accessed September 22, 2025, https://www.researchgate.net/publication/263311175_Air_quality_modelling_simulation_and_computational_methods_a_review
48. A gridded air quality forecast through fusing site-available machine learning predictions from RFSML v1.0 and chemical transport model results from GEOS-Chem v13.1.0 using the ensemble Kalman filter - GMD, accessed September 22, 2025, https://gmd.copernicus.org/articles/16/4867/
49. FastCTM (v1.0): Atmospheric chemical transport modelling with a principle-informed neural network for air quality simulations - GMD, accessed September 22, 2025, https://gmd.copernicus.org/preprints/gmd-2024-198/
50. Fusing Observational, Satellite Remote Sensing and Air Quality Model Simulated Data to Estimate Spatiotemporal Variations of PM 2.5 Exposure in China - MDPI, accessed September 22, 2025, https://www.mdpi.com/2072-4292/9/3/221
51. Assessing the Impact of Wildfire Smoke Transport Through Chemical Transport Modeling, Satellite Retrievals, and Ground-Based Observations of Ozone in Rural Nevada - PubMed Central, accessed September 22, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC12261271/
52. Development of a parametrised atmospheric NOx chemistry scheme to help quantify fossil fuel CO2 emission estimates - EGUsphere, accessed September 22, 2025, https://egusphere.copernicus.org/preprints/2025/egusphere-2024-3949/egusphere-2024-3949.pdf
53. (PDF) A Hybrid Model for Air Quality Prediction Based on Data ..., accessed September 22, 2025, https://www.researchgate.net/publication/351628307_A_Hybrid_Model_for_Air_Quality_Prediction_Based_on_Data_Decomposition
54. Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea - ACP, accessed September 22, 2025, https://acp.copernicus.org/articles/19/1097/2019/

55. Using Satellite Data for Air Quality and Health Applications | Vital Strategies, accessed September 22, 2025, https://www.vitalstrategies.org/wp-content/uploads/Using-Satellite-Data-for-Air-Quality-and-Health-Applications.pdf
56. Full article: Spatially continuous mapping of hourly ground ozone levels assisted by Himawari-8 short wave radiation products - Taylor & Francis Online, accessed September 22, 2025, https://www.tandfonline.com/doi/full/10.1080/15481603.2023.2174280
57. Research progress, challenges, and prospects of PM2.5 ..., accessed September 22, 2025, https://cdnsciencepub.com/doi/10.1139/er-2022-0125
58. Air Pollution Trends and Predictive Modeling for Three Cities with Different Characteristics Using Sentinel-5 Satellite Data and Deep Learning - MDPI, accessed September 22, 2025, https://www.mdpi.com/2073-4433/16/2/211
59. Evaluation Metrics for Air Quality Optimization Utilizing Machine Learning: PRISMA Review, accessed September 22, 2025, https://www.researchgate.net/publication/390008799_Evaluation_Metrics_for_Air_Quality_Optimization_Utilizing_Machine_Learning_PRISMA_Review
60. A 6-year-long (2013–2018) high-resolution air quality reanalysis dataset in China based on the assimilation of surface observations from CNEMC - ESSD Copernicus, accessed September 22, 2025, https://essd.copernicus.org/articles/13/529/2021/
61. A standardized methodology for the validation of air quality forecast applications (F-MQO): lessons learnt from its application across Europe - GMD, accessed September 22, 2025, https://gmd.copernicus.org/articles/16/6029/2023/
62. Evaluating the Performance of Air Quality Models - DEFRA UK Air, accessed September 22, 2025, https://uk-air.defra.gov.uk/reports/cat05/1006241607_100608_MIP_Final_Version.pdf
63. From EarthData to Action: Cloud Computing with Earth Observation Data for Predicting Cleaner, Safer Skies - NASA Space Apps Challenge, accessed September 22, 2025, https://www.spaceappschallenge.org/2025/challenges/from-earthdata-to-action-cloud-computing-with-earth-observation-data-for-predicting-cleaner-safer-skies/
64. New Satellite-Driven Model Tracks Long-Term Air Pollution Trends Across China, accessed September 22, 2025, https://sph.uth.edu/news/story/new-satellite-driven-model-tracks-long-term-air-pollution-trends-across-china
65. Model Forecast | Air Quality - NASA, accessed September 22, 2025, https://airquality.gsfc.nasa.gov/forecast