Frontiers of Environmental Science & Engineering
2025, 19(12): 172 https://doi.org/10.1007/s11783-025-2092-6

REVIEW ARTICLE

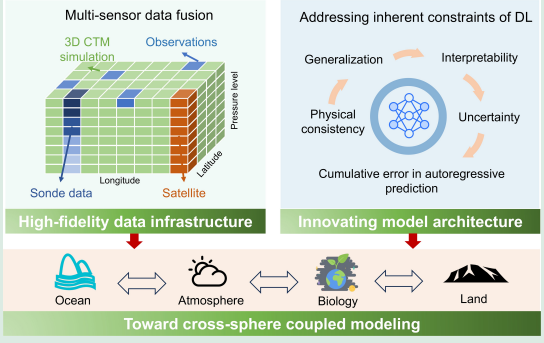
Deep learning for air pollutant forecasting: opportunities, challenges, and future directions

Chenliang Tao[1], Yiheng Wang[1], Yuhao Wang[1], Zhonghua Zheng[2], Hongliang Zhang ✉ [1,3,4]

1. Department of Environmental Science and Engineering, Fudan University, Shanghai 200438, China
2. Department of Earth and Environmental Sciences, The University of Manchester, Manchester M13 9PL, UK
3. School of Environment and Architecture, University of Shanghai for Science and Technology, Shanghai 200438, China
4. IRDR ICoE on Risk Interconnectivity and Governance on Weather/Climate Extremes Impact and Public Health, Fudan University, Shanghai 200438, China

HIGHLIGHTS

- Real-time data assimilation is essential to build reliable and actionable forecasting tools.
- Advances in model architectures help to overcome the limitations of deep learning-based air quality forecasting.
- Cross-sphere coupling hold promise for enabling data-driven seasonal air pollutant forecasting.



ABSTRACT: Deep learning methods are increasingly employed to forecast air quality from an ever-increasing stream of data and algorithms. However, the efficacy of current approaches may be questionable when evaluated not solely in terms of greater forecasting fidelity, but also concerning the decision-making process in pollution early warning. Here, rather than amending classical machine learning algorithms, we argue that now is the time to push the frontiers of air pollutant forecasting beyond state-of-the-art approaches. This can be achieved through near real-time assimilation of multi-scale observations for laying the foundation of training data, enhanced attribution methods for impending heavy pollution, diagnostics for forecasting uncertainty, and advanced climate-chemistry emulators for improving seasonal forecasting. To harness this potential, it is essential to address several key challenges in deep learning methods, particularly generalization ability in extreme events, physics-informed interpretable approaches, and the mitigation technology of cumulative errors in multi-process coupled systems. This interdisciplinary endeavor will remain a central pursuit in the quest to anticipate and manage environmental change.

KEYWORDS: Deep learning, Air pollution forecasting, Data assimilation, Seasonal forecasting

✉ Corresponding author. E-mail: zhanghl@fudan.edu.cn

Article history: Received 25 May 2025, Revised 3 August 2025, Accepted 8 September 2025, Available online 30 September 2025

© The Author(s) 2025. This article is published with open access at link.springer.com and journal.hep.com.cn

Special Issue—Atmospheric and Earth System Modeling towards Coordinated Pollution Control and Climate Change Mitigation (Guest Editors: Yuhang Wang, Hongliang Zhang, Yang Gao, Bin Zhao & Peng Wang)

1 Introduction

Environmental challenges have escalated to the forefront of global concerns, with issues like air pollution, climate change, and biodiversity loss posing significant threats to public health and ecological stability (Hong et al., 2019; Chen et al., 2023). Among these, air pollution released into the atmosphere and transformed by anthropogenic activities, volcanic eruptions, forest fire releases, and natural sources contributes to millions of premature deaths annually and affects the quality of life for billions worldwide (Cohen et al., 2017). Accurate forecasting of air pollutant concentrations, such as particulate matter smaller than 2.5 micrometers ($PM_{2.5}$), ozone (O_3), and nitrogen dioxide (NO_2) is essential for early warning, mitigation strategies, and informed policy-making. A success story in air pollutant short-term forecasting is process-based chemical transport models (CTMs), which have been improved by the encoding of detailed atmospheric chemistry and transport processes and the assimilation of established observation systems (Appel et al., 2021). Climate-chemistry models have laid the foundation for seasonal predictions by coupling the physical climate system with the chemical processes that govern atmospheric composition (Caldwell et al., 2019). Nevertheless, predictions of future air quality evolution remain contingent upon sufficiently accurate emission inventories and chemical regimes (Byun and Schere, 2006; Lamarque et al., 2012), and often struggle with high computational costs, particularly when attempting to achieve finer temporal and spatial resolutions.

Recently, deep learning (DL) has emerged as a powerful data-driven paradigm that can extract interpretable knowledge representations from a deluge of available earth system data with increased transmission rates exceeding hundreds of terabytes per day (Agapiou, 2017; Reichstein et al., 2019). In fields like computer vision and sequence modeling (Dosovitskiy et al., 2021; Liu et al., 2021a), deep neural networks have achieved remarkable successes by automatically learning rich feature representations. This success has spurred a growing interest in applying DL techniques to model the highly non-linear, multi-scale spatiotemporal characteristics of meteorological dynamics for weather forecasting, as a way to break through the limits of tradi-tional numerical models (Bi et al., 2023; Lam et al., 2023; Kochkov et al., 2024; Price et al., 2025).

Unprecedented data sources, including model simulations such as the Climate Model Intercomparison Project (CMIP), satellite remote sensing, as well as comprehensive ground monitoring networks (Bucsela et al., 2013; Collins et al., 2017; Copernicus Sentinel-5P, 2020; Lyapustin and Wang, 2022), in conjunction with the advent of advanced computational capabilities (e.g., high-performance GPUs) and advanced neural network architectures, including Transformer models and Diffusion models (Vaswani et al., 2017; Ho et al., 2020; Karras et al., 2022), have unlocked exciting opportunities to advance our understanding of atmospheric environmental sys-tems from data (Zhang et al., 2025). However, these tools require further development and adaptation to meet the specific needs of atmospheric research, which involves interactions between various physical, chemical, and biological processes and meteorological conditions over varying scales. Another major task in the coming broader adoption of DL is deriving models that can understand the underlying processes driving their predictions, promoting our evolving understanding of the physical laws governing atmospheric processes, rather than being merely powerful models that operate as “black boxes”.

We aim to explore the cutting-edge applications of DL in the field of air pollution forecasting, examining both its potential and challenges. We will highlight the key areas where DL can advance the field, including predictive accuracy, data assimilation, and the modeling of complex spatiotemporal patterns. We will also discuss the integration of domain knowledge to address issues of interpretability, physical consistency, and uncertainty, ultimately paving the way for more robust and reliable environmental prediction models. By embracing these new opportunities, DL has the potential to revolutionize our understanding of air pollution and its broader environmental impacts, offering the ability to forecast pollution levels with greater precision and responsiveness. This paper will provide a comprehensive overview of the state-of-the-art methods, challenges, and future directions for DL in the context of air quality forecasting and beyond.

2 Progress in deep learning for air quality forecasting

The advent of DL technology has precipitated a paradigm shift in the domain of air pollutant concentration prediction, owing to its advanced nonlinear modeling capabilities and adaptable architectural design (Ravi et al., 2024; Fan et al., 2025). Traditional machine learning has faced limitations in handling the complexity of air pollutant concentration prediction.

They often rely on predefined, manual feature engineering and struggle to capture the intricate spatial and temporal dependencies inherent. In contrast, DL can automatically extract high-dimensional spatial and temporal dependencies from heterogeneous data from multiple sources, such as satellite remote sensing, ground monitoring stations, and meteorological simulations (Lu et al., 2021; Huang et al., 2022). By establishing direct mapping relationships between input variables and pollutant concentrations, DL can realize cross-scale modeling and forecasting of complex atmospheric systems. For instance, traditional machine learning methods for remote sensing image segmentation rely on shallow features, such as texture and color gradients, limiting their ability to capture high-level semantic information (Watanachaturaporn et al., 2006; Li et al., 2024a). DL excels in multi-scale feature extraction and integration, significantly outperforming traditional models (Bai et al., 2022). This hierarchical abstraction capability enables the adaptive fusion of multi-scale information to avoid the limitations of manually preset feature scales.

The application of machine-learning algorithms for air pollution forecasting may be traced back to the initial prediction of future pollutant concentrations from continuous time-series datasets obtained from fixed observation stations. Zhang et al. (2023) utilized a Transformer model that employs a multi-head sparse attention mechanism to predict future pollutant concentrations at air quality monitoring stations using correlation information among the entire $PM_{2.5}$ time-series data captured. However, the inherent limitation of the lack of continuous spatial coverage of air quality station forecasts restricts their ability to perform compre-hensive environmental assessments. This paucity of spatial resolution has facilitated the development of seamless air quality forecasting. For instance, Bi et al. (2022) utilized the Random Forests model to predict high-resolution $PM_{2.5}$ concentrations for the subsequent five days, drawing upon CTM simulation data, thereby ensuring near-real-time air quality forecasting under limited computational resources. An alternative effective forecasting approach involves the integration of deep-learning models with CTMs, wherein researchers construct residual neural network simulators to replace the gas-phase chemistry module of the global nested grid air quality forecasting system. The model has been shown to exhibit a processing speed that is 300–750 times faster than existing models while maintaining simulation accuracy (Wang et al., 2022). Recent research has begun experimenting with an end-to-end learning framework that enables direct mapping from historical pollutant distributions, meteorology, and precursor emissions to changes in air pollution over the next few days, independent of coupling with the CTM, thereby improving forecasting efficiency and accuracy (Lyu et al., 2024; Hu et al., 2025). For instance, a foundation model trained on over one million hours of diverse geophysical data surpasses operational forecasts in air quality prediction, underscoring the transformative impact of DL on environmental forecasting (Bodnar et al., 2025).

3 High-quality data scarcity

The parallels between DL applications in weather and climate forecasting and air pollution forecasting applications outlined above are striking. Nevertheless, a plethora of discrepancies remain. Chief among these issues is the limitation of data quality, which is distinct from the field of weather forecasting, which benefits from high-quality reanalysis datasets such as ECMWF's ERA5. The air quality model is deficient in the absence of a comprehensive, long-term, and mechanism-consistent dataset of pollutant simulation or reanalysis data. This is due to the substantial yet elusive influence of anthropogenic emissions on air quality, which are highly variable and difficult to quantify or track, and the complexity of interacting mechanisms when pollutants enter the atmosphere (Zhao et al., 2025). For example, the atmosphere-only climate models with prescribed sea surface temperature would fail to capture the seasonally evolving teleconnections of surface O_3 and the Atlantic Multidecadal Oscillation (Shen and Mickley, 2017). It is important to note that emissions data may be outdated or uncertain, especially in regions without rigorous inventory reporting. For example, a newly constructed factory or power plant may emit pollutants for months or even years before this is reflected in emission databases, resulting in a data-driven model that remains unaware of this novel source of pollution.

The integration of simulations with ground monitors and satellites has been demonstrated to facilitate the resolution of discrepancies in data quality (Requia et al., 2020; Geng et al., 2021; Huang et al., 2021a; Li et al., 2022b; Thongthammachart et al., 2022). This is a complex task for reaching a level that can support the training of air quality forecasting models, taking into account the inherent differences in spatiotemporal resolution, physical mechanism, content, and vertical profiles of the data products. Monitoring station networks are highly heterogeneous, with significant portions of the Global North and urban areas, resulting in vast rural areas chronically under-monitored even in

data-abundant countries (Wei et al., 2023). Models trained predominantly on datasets assimilated by urban observations face inherent limita-tions when applied to regions with distinct pollution signatures. Furthermore, unrecorded pollution events, such as wildfires or dust storms in data-sparse deserts, remain entirely absent from the training signal, precluding any possibility of predictive learning in those contexts.

Satellite data partially mitigates these gaps but introduces its challenges. Satellite observations are inherently discontinuous due to their reliance on orbital dynamics, in which the polar-orbiting satellite might provide one observation at a specific moment per day (or even fewer) for a given location, and these observations are highly susceptible to cloud interference. As a result, atmospheric measurements are often missing in 60%–70% of global scenes (Christopher and Gupta, 2010), leading to substantial data gaps. Heavy pollution episodes can coincide with weather conditions that either form clouds or with dense smoke, leading to a systematic bias toward clear skies and certain seasons, skewing the training data that a DL model sees. Researchers have had to devise gap-filling methods to create "gapless" datasets, but any such interpolations or model-based fill-ins carry their uncertainties (Hammer et al., 2023; Bai et al., 2024).

Another major challenge of integrating satellite data to compensate for sparse ground-based concerns is the approach development. Initially, land use regression models were employed to estimate pollutant concentrations by correlating ground-level measure-ments with land use characteristics and satellite column density (Larkin et al., 2023). With advancements in machine learning, ensemble methods like Random Forest and XGBoost have been utilized to enhance spatiotemporal resolution and accuracy in estimating surface-level pollutants such as $PM_{2.5}$ (Liu et al., 2022; Zhu et al., 2022; Zheng et al., 2023; Chen et al., 2024a). The efficacy of these models has been demonstrated in several studies, thus indicating that they offer enhanced predictive capabilities in comparison to traditional statistical approaches. However, a fundamental limitation of these models lies in their treatment of spatial and temporal dependencies. They establish associations between satellite-derived pollutant column density and ground-based measurements in specific grid cells without explicitly modeling the spatiotemporal autocorrelation inherent in atmospheric processes, resulting in a lack of robustness in capturing the dynamics of pollutant dispersion and transformation over time and space. Consequently, their predictive accuracy diminishes in regions with sparse observational data, where understanding spatiotemporal patterns is crucial.To address these, while recent research has focused on developing methods that incorporate spatiotemporal weighting mechanisms to capture the impact of the behavior in neighboring pollutants to local variations (Di et al., 2016; Wei et al., 2022; Tao et al., 2024b), there is still no fundamental solution to the limitations of full-field reconstruction of spatiotemporal dynamics.

Reliable emission inventories are pivotal for air quality forecasting. Traditional bottom-up estimations of air pollutant emissions based on activity data face significant limitations due to the delayed availability of activity statistics (Lu et al., 2020). In contrast, satellite-based top-down methods offer near-real-time constraints on emissions. CTM-based inversion models are commonly employed to improve emission estimates (Miyazaki et al., 2012). However, the challenge with these methods lies in the precise and efficient quantification of the relationship between ambient air pollutant concentrations and emissions. DL has proven to be effective in capturing the nonlinear emission-concentration response. These methods successfully adjust emissions by backpropagating gradients from the loss function, which quantifies the discrepancy between CTM predictions and observations (Huang et al., 2021b; He et al., 2022b). Xing et al. (2022) further advanced this approach by integrating a variational autoencoder (VAE) encoder (inverse model) for emission optimization with a VAE decoder for updating concentration changes based on emission variations, enabling posterior emission estimation. Furthermore, Li and Xing (2025) demonstrated that coupling a DL-driven observation-based emission optimization module with a forecasting module led to significant improvements in the first 24-h forecasting accuracy. These findings underscore the importance of initial emission and concentration field corrections. Despite these advancements, challenges remain in achieving dynamic optimization of emission inventories with high temporal and spatial resolution, particularly when focusing on sector-specific emissions.

To address persistent gaps and biases in atmospheric observations and simulations, a new paradigm that leverages generative DL models to assimilate heterogeneous observations across spatial, temporal, and vertical dimensions in data fusion is emerging. Specifically, a flexible, three-dimensional data assimilation system is necessary to integrate Polar-orbiting satellites for broad spatial coverage with daily column density (Copernicus Sentinel-5P, 2020), geostationary satellites for filling in temporal gaps to capture diurnal variability (Song et al., 2022), ground-level measurements for accurate near-surface conditions, and sonde data to constrain chemical vertical profiles across

tropospheric and strat-ospheric layers (Fig. 1) (Wang et al., 2018; Liu and Xing, 2024). Critically, advances in generative models, such as masked autoencoders and diffusion-based archi-tectures, offer unprecedented opportunities to impose learned, data-driven constraints on chemical transport model (CTM) outputs (He et al., 2022a; Li et al., 2024b; Price et al., 2025). These models can reconstruct spatiotemporal dynamics from sparse measurements by learning from the joint distribution of multisource inputs to flexibly reconcile observational inconsi-stencies and improve the realism of model-derived pollutant fields. This approach represents a shift from deterministic assimilation toward probabilistic, gene-rative fusion, where uncertainty cannot merely be managed but modeled.

4 Pitfalls of deep learning

4.1 Generalization ability in extreme conditions

Despite the great progress, one of the most critical and challenging aspects of air quality forecasting is the extreme conditions that pose great potential risks to human health, such as severe pollution episodes or unusual weather patterns (Mo et al., 2022; Tao et al., 2024a). These rare events are underrepresented in training data, leading to imbalanced learning where models are biased toward moderate conditions (Liu et al., 2021b; Ran et al., 2025). As a result, predicting the high concentration "spikes" of pollutants remains difficult, in which heavy pollution days containing concentration maxima are a stringent test of model robustness. Even advanced DL architectures can see sharp drops in accuracy when forecasting extreme events or operating outside the range of their training data. Furthermore, models trained in one city or season often fail when applied to a different region or climate, reflecting poor transferability across spatiotemporal domains (Zhang, 2025). By affecting the probability and intensity of extreme events (wildfires and heat waves) (Libonati et al., 2022; Xie et al., 2022), global warming may cause DL models trained in the past decades to systematically underestimate future extreme events. This limited generalization indicates that without special measures, purely data-driven predictors will remain unreliable exactly when high-stakes extreme air pollution events occur.

Recent advances aim to broaden the adaptability of DL to handle extreme conditions. Transfer learning has shown particular promise. Models pre-trained on large datasets can be fine-tuned on a target city with limited data, dramatically improving generalization to new conditions (Ma et al., 2020; Jairi et al., 2024; Bodnar et al., 2025; Jiang et al., 2025). Ensemble modeling can further enhance robustness in extremes by combining predictions from multiple models, each with different architectures or trained on different data subsets (Liu et al., 2019; Guastavino et al., 2022; Tao et al., 2024a). The ensemble can capture a wider range of behavior

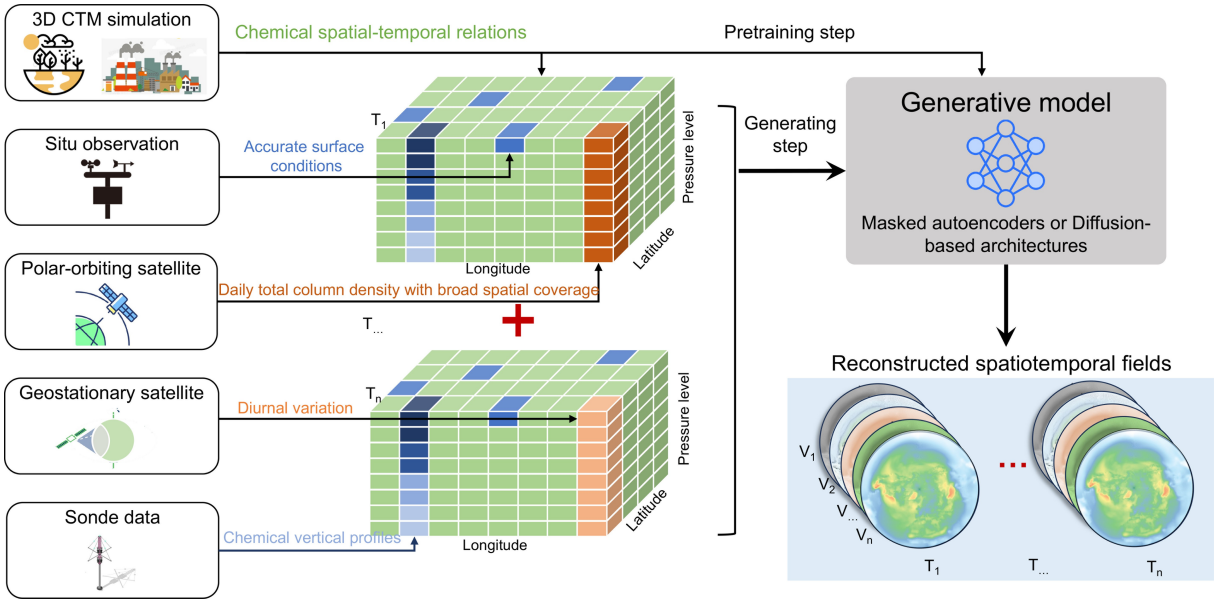


Fig. 1 Structure of the multi-sensor data fusion utilizing deep learning. The cube illustrates the spatial dimension of the data, while the blocks represent the grid corresponding to specific altitudes, latitudes, and longitudes.

and avoid the failure of any single model on an out-of-distribution event. Additionally, methods such as injecting synthetic extreme events or perturbing inputs for data augmentation, bias corrections, and methods similar to meteorological ensemble forecasting can ensure rare high-pollution scenarios are adequately considered (Ehrendorfer, 1997; Zhang et al., 2023). Whittaker and Di Luca (2025) predicted the initial conditions that can accurately forecast extreme heatwaves by minimizing the loss function of the intensity of temperature abnormality and the disturbance amplitude of the initial conditions. Then, a more extreme heatwave scenario was generated than the traditional ensemble simulation, effectively reproducing the temperature distribution of ERA5 data, especially at the extreme temperature values of 95th and 99th percentiles. Together, these approaches improve the resilience of DL forecasting, ensuring that models maintain reasonable accuracy even under extreme conditions.

4.2 Interpretability and physical consistency

While enhancing forecast accuracy remains the paramount objective, this is insufficient on its own. To attain optimal application of DL models, it is essential that they are both reliable and interpretable. Reliability is grounded in the established physicochemical laws that govern the atmospheric system, while interpretability facilitates the identification of areas for model enhancement and bolsters understanding of the process being modelled (Montavon et al., 2018). Another reason why explainability and physical consistency are so important in the air quality forecasting is that domain experts still need to understand in advance which factors (e.g., expected emissions, temperature, wind-induced transportation) will drive future changes in air pollutants to carry out emission control in advance and avoid the occurrence of heavy pollution events that harm human health, while ensuring that the predictive behavior of the model is consistent with the principle of atmospheric science. Explainable DL attempts to capture complex and unresolved knowledge in a way that traditional empirical formulas cannot. Early researches show that, when carefully trained, this method can discover widely verified atmospheric chemical knowledge and even reveal new relationships in the data (Zhan et al., 2022; Tao et al., 2024b). Considerable progress has been made in the field of traditional machine learning interpretation methods, with the SHAP interpretable method being a notable example of this advancement. This method has enhanced the ability to mine observed laws from data. However, there has been relatively little progress in the area of black-box DL, which is characterized by its high level of complexity, hindering our ability to understand the reasoning behind predictions (Zhang, 2025). The lack of transparency also makes it difficult to trust model outputs in high-stakes scenarios.

To increase trust and practicality, a straightforward approach is to apply post-hoc explainability methods borrowed from explainable AI (XAI) (Lundberg and Lee, 2017; Molnar, 2025). This method helps to determine the driving factors on which the model depends. Attention maps and Grad-CAM can also highlight which time steps or spatial locations the neural network focuses on for a particular forecast (Abnar and Zuidema, 2020; Selvaraju et al., 2020). In addition, a promising frontier is physics-informed DL, where domain knowledge is embedded into the model structure or training objective to ground the model in real physical processes. A recent example is a physics-guided neural network that integrates the diffusion and advection equations, which govern the processes of pollutant transport, directly into its architecture (Hettige et al., 2024). The model was proven to capture underlying physical processes of particle movement and generate accurate predictions with real physical meaning. Similarly, hybrid models that combine a CTM with a neural network have been shown to maintain better consistency with known science. These physics-based or hybrid models have dual advantages. They improve transparency by aligning model computations with scientific concepts and even enhance generalization due to physical laws acting as a regularized form in the online simulation (Kelp et al., 2020, 2022; Xu et al., 2022). Moreover, the efficient characteristics of DL computing support a large number of sensitivity experiments by permuted input features (e.g., emission), and to a certain extent, increase our understanding of how emission changes affect the causes of atmospheric pollution (Xing et al., 2020). As research in interpretable DL progresses, we are seeing the gap between complex neural nets and human-understandable models narrow. The community is moving toward "opening up" the black box, ensuring that DL models for air quality can be scrutinized and trusted much like traditional models, without sacrificing their superior performance.

4.3 Uncertainty quantification

Weather forecasting is inherently uncertain due to the nonlinear nature of atmospheric dynamics and its intrinsic sensitivity to initial conditions, whereby even

small perturbations can lead to large and growing deviations in the predicted evolution (Ehrendorfer, 1997; Sha et al., 2024). Two forecasts with the same mean value may have very different confidence levels depending on recent weather dynamics. Modern weather forecasting systems address this issue by employing ensemble methods, typically generating 30–50 forecasting members through slight perturbations in initial conditions or physical model parameters, thus sampling a range of plausible future states (Buizza, 2019). However, in current operational air quality forecasting systems, the impacts of meteorological uncertainty on pollutant predictions are not explicitly quantified using ensemble approaches. This gap largely stems from the prohibitive computational cost of running three-dimensional CTM multiple times with perturbed meteorological inputs, rendering ensemble-based air quality forecasting impractical for daily operational use.

Standard deep-learning forecasting provides point estimates of pollutant concentrations without any indication of uncertainty. Yet most data-driven solutions lack proper quantification of model uncertainty to communicate how much to trust a given prediction. In air quality management, the inability to quantify uncertainty can result in the propagation of false security, resulting in decision-makers being unable to distinguish a highly reliable forecast from one that is conjectural., The challenge is further compounded by multiple sources of uncertainty. Aleatoric uncertainty stems from random fluctuations in emissions and turbulent transport, which limit the predictability even with perfect modeling. Epistemic uncertainty stems from the limited knowledge in the model, for instance, encountering a weather and air pollution regime not seen during the training process. Traditional DL models often capture neither type well. In the absence of explicit uncertainty estimation, a "black-box" air quality model may exhibit over-confidence in regions characterized by sparse data or during extreme events, thereby diminishing its utility for risk-averse decision processes.

Embedding the ensemble forecasting framework into the air quality DL model is a potential solution. One approach is Bayesian DL, which treats model parameters (weights) as distributions rather than fixed values. Bayesian Neural Networks inherently quantify epistemic uncertainty by learning posterior over-weights, yielding predictive distributions instead of single values (Franchi et al., 2024). Another practical alternative has gained traction by training multiple neural networks with different initializations or training subsets and combining these outputs (Ren et al., 2022; Zhang et al., 2023). In this approach, the differences between the ensemble members are treated as an empirical measure of model uncertainty. Similarly, Monte Carlo Dropout entails randomly dropping neurons during prediction in a repeated fashion, effectively generating an ensemble from one network to infer a confidence interval (Gal and Ghahramani, 2016). Recent studies have demonstrated that probabilistic DL models are more appropriate for decision-making processes, as they are capable of conveying forecasting confidence and have the potential to minimize false alarms (Chen et al., 2024b; Schreck et al., 2024). In the future, it is anticipated that uncertainty quantification will become a standard expectation for advanced forecasting models, on par with accuracy, to ensure reliable use in practice.

4.4 Cumulative error in autoregressive prediction

A persistent difficulty in time-series forecasting is that small prediction errors can snowball over multiple time steps. In air quality forecasting, a model that predicts hour-by-hour or day-by-day concentrations may feed its predictions back as inputs for the next time step, a procedure that can cause compounding errors. If the model is slightly off at first, the mistake may be amplified at each subsequent step, leading to rapidly diverging forecasts. Without intervention, iterative DL forecasting tends to drift from reality, sometimes predicting implausible trajectories after several steps. Closely related is the issue of physical consistency. Purely data-driven models do not inherently enforce the known physical laws and constraints of atmospheric chemistry. Consequently, they might produce outputs that violate basic principles.

To address the issue of error accumulation, it is necessary to implement more sophisticated training techniques and to incorporate physical knowledge. The employment of regularization, like dropout and L2 weight penalties and modified training strategies, has emerged as a pivotal approach to mitigate the propagation of errors (Zhou et al., 2019). Rather than training a model exclusively to predict the next time step, one can train it to handle multi-step prediction sequences by exposing it to its predictions during the training process (Chen et al., 2025). This approach enables the model to self-correct its errors. The use of an ensemble of models trained for various lead times has also been demonstrated to enhance the accuracy and robustness of predictions. This hierarchical temporal aggregation approach, as shown in the Pangu-Weather

model, helps in reducing the propagation of errors across multiple steps by greatly reducing the number of iterations (Bi et al., 2023; Lin, 2024).

The imposition of constraints on the model outputs utilizing the design of the loss function can penalize physically impossible results (Beucler et al., 2021). For instance, it can be used to discourage the model from ever predicting negative concentrations (Shen et al., 2024). Beyond hard constraints, the coupling of DL models with CTM or Earth system models is a powerful way to inject physical realism. In this framework, the DL model can correct biases and capture complex patterns, while the CTM provides physical guidance and baseline consistency, thereby achieving higher skill and extended range than was previously possible (Fig. 2) (Kelp et al., 2022). It is asserted that the physical information architecture will contribute to the stability and consistency of prediction. Researchers have achieved stable autoregressive rollouts for a year of simulated time (1460 steps) in forecasting atmospheric dynamics by introducing Spherical Fourier Neural Operators to learning operators on spherical geometries (Bonev et al., 2023). This development has the potential to significantly enhance the performance and scientific coherence of next-generation air quality models, ensuring reliable guidance even over extended forecasting horizons.

5 Difficulties in the seasonal modeling

While DL models have shown considerable skill in short-term air quality forecasting, they encounter significant challenges in extending their capabilities to longer-term forecasting and simulation. Seasonal variations in air pollutant concentrations are strongly influenced by climatic factors, making it necessary first to understand and predict how the climate forcing and meteorological patterns deviate from normal. Short-term weather and air pollutant forecasting is primarily an initial condition problem, where the future atmospheric state is largely determined by preceding states, rendering autoregressive models highly effective for such tasks (Fig. 3). In contrast, seasonal forecasting is fundamentally a boundary condition problem, wherein the future climate state is driven more by interactions between the atmosphere and its boundaries, such as land surfaces and oceans (Mariotti et al., 2018). Within this framework, internal climate variability, such as multi-decadal oscillations representing atmospheric circulation patterns or coupled atmosphere-ocean systems (Mechoso, 2020), can cause ebbs and flows in global temperature or regional climate that temporarily mask or enhance the effects of external forcing. However, both initial and boundary conditions offer limited predictability, resulting in what is often referred to as a "predictability desert" for seasonal forecasting (Domeisen et al., 2022). If DL is applied in isolation, it lacks the necessary foresight to capture meteorological variability over extended periods. More critically, errors inherent in seasonal climate predictions can propagate and accumulate, leading to exponentially increasing uncer-tainties in air quality forecasting as the lead time extends.

Expanding air quality forecasting to seasonal

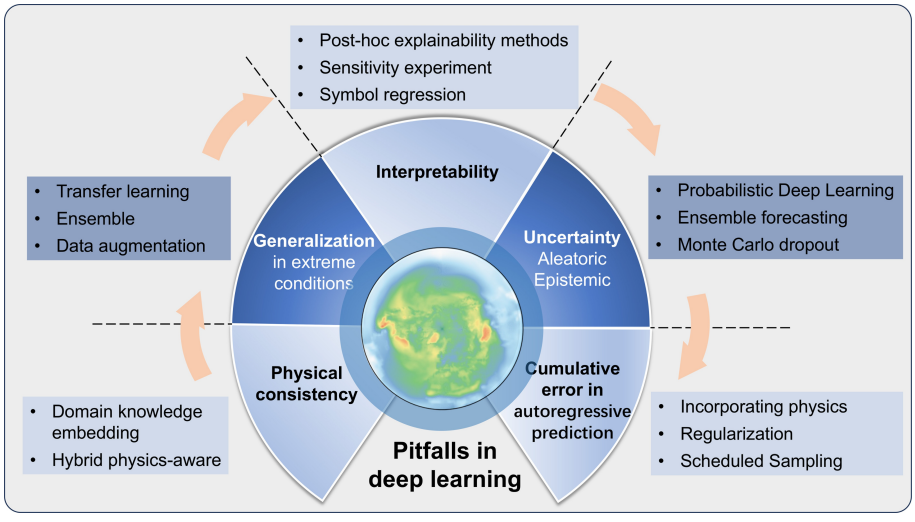


Fig. 2 Key challenges and emerging trends in the application of deep learning techniques for air pollution forecasting. Despite notable advancements, the full potential of deep learning in air pollution forecasting has yet to be fully realized.

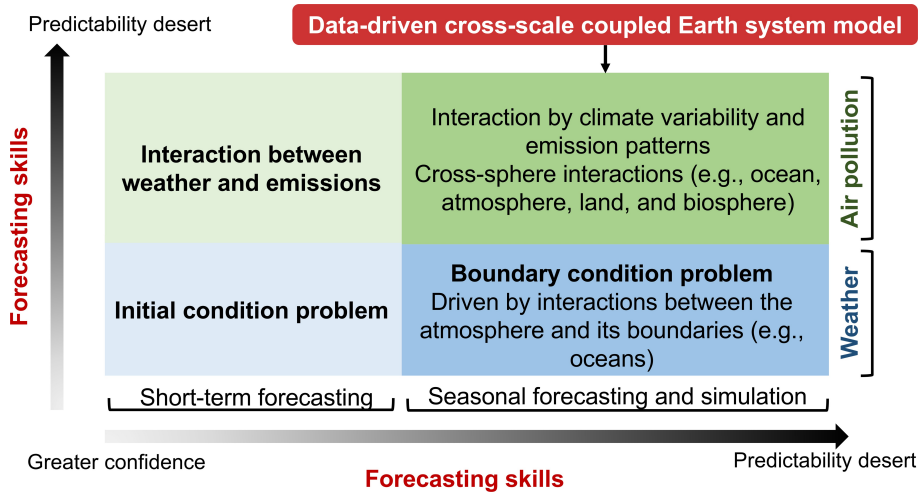


Fig. 3 Comparison of deep learning forecasting skills and potential directions in short-term and seasonal forecasting.

timescales requires accounting for trends in climate-driven shifts such as atmospheric stagnation and heat-waves, emissions regulations, land use changes, and emission patterns (Baklanov et al., 2017). On the one hand, climate variability can significantly influence air quality. For example, the Atlantic Multidecadal Oscillation has been shown to drive multidecadal variability in surface ozone concentrations over the United States (Shen and Mickley, 2017). On the other hand, atmospheric pollutants can perturb the energy balance of the atmosphere. Aerosols from wildfire produce an instantaneous radiative forcing of -1.0 ± 0.6 W/m^2 over cloud-free ocean regions by directly scattering and absorbing solar radiation, thereby exerting a notable cooling effect (Hirsch and Koren, 2021). In addition, aerosols serve as cloud condensation nuclei, altering cloud albedo and further amplifying perturbations to the surface atmosphere energy balance (Li et al., 2022a; Zhong et al., 2024; Xiong et al., 2025). The extent to which such climate and chemistry interactions are represented within DL models constitutes a critical factor limiting the accuracy of long-term air quality predictions, and addressing this challenge remains an important direction for future methodological development.

The recognition that physical climate systems and atmospheric composition are deeply interlinked has driven a rapid evolution from siloed modeling approaches toward integrated models that couple the ocean-atmosphere-land-biosphere system with atmospheric chemistry. This goal is to simulate the Earth system as a whole, capturing critical feedback loops in a physically consistent way. One fundamental challenge lies in modeling cross-sphere interactions with sufficient fidelity. Because these models span the ocean, atmosphere, land, and biosphere, errors or gaps in one component can cascade into others. Many studies used one-way off-line methods, where atmospheric chemistry was simulated using prescribed meteorological data (Zhu et al., 2023; Lyu et al., 2024). By contrast, online-coupled models or chemistry-climate models simulate meteorology and chemistry simultaneously within the same system, enabling feedback between the two (Baklanov et al., 2014). For example, an online model can account for how a surge in aerosol pollution might reduce solar radiation and stabilize the atmosphere, which in turn could alter wind patterns and pollutant dispersion, processes that a one-way model would miss (Baró et al., 2017). These reflect the fact that phenomena like “chemical weather” and climate cannot be reliably understood in isolation (Griffiths et al., 2024). Coupling different model components also means dealing with disparate spatial and temporal scales: the atmosphere responds on timescales of hours to days, while the ocean and biosphere might adjust over years or decades. Aligning these in one framework without introducing artificial shocks or drift is difficult.

The rise of artificial intelligence is opening new frontiers in coupled model development. Recent advances have demonstrated that DL models, leveraging their high computational efficiency, can generate large ensembles within limited computational budgets and outperform leading subseasonal-to-seasonal (S2S) forecasting systems, such as the ECMWF model, in predicting total precipitation and outgoing longwave radiation (Chen et al., 2024c). These highlight the ability of DL to extract and leverage subtle long-range dependencies, thereby offering improved predictive skill at seasonal

timescales. DL-based simulations have successfully reproduced key features of atmosphere-ocean coupling, including the propagation of tropical oceanic waves with realistic phase speeds and the internal generation of El Niño–Southern Oscillation (ENSO) events (Wang et al., 2024). This consitutes a prototypical examplar, exhibiting the shared characteristics inherent in the challenge of multi-spheric interactions in forecasting. Additionally, hierarchical multi-task frameworks could provide an effective way to model the interactions across multiple spheres by the shared representations across these tasks. Building on these successes, new opportunities emerge for long-term air quality modeling, suggesting that data-driven Earth system models, potentially in the form of digital twins, could better integrate atmosphere, ocean, land, and chemistry, with continuous refinement through observational data assimilation.

6 Conclusions

DL applied to air pollution prediction has increasingly become a research hotspot in the field of atmospheric environment, marking a breakthrough in the interdisciplinary field of environmental science and artificial intelligence. DL is gradually superseding the inherent limitations of traditional methods based on the characteristics of automatic feature extraction, end-to-end prediction, spatiotemporal modeling, and uncertainty quantification, providing unprecedented technical support for pollution early warning. Nevertheless, only when the above challenges are effectively addressed, DL is poised to become an increasingly indispensable part of air quality forecasting.

Future research will likely center on hybrid approaches that blend data-driven algorithms with domain knowledge, yielding models that are accurate, interpretable, and physically grounded. Efforts to improve generalization suggest that integrating real-time data assimilation, in which DL models quickly adjust forecasts based on the latest observations, will be key for ensuring both the accuracy and adaptability of the model and handling new scenarios and extreme events. It can provide very strong theoretical constraints by imparting knowledge of the dominant physical rules of the Earth system from observations to models. Interpretability and explainability will be prioritized not only through the utilization of XAI tools but also via the design of inherently interpretable architectures and physics-informed neural networks that satisfy scientific consistency by construction. Likewise, uncertainty quantification is anticipated to transition into the mainstream of operational forecasting, with probabilistic forecasts becoming the prevailing outputs alongside deterministic estimates.

Another emerging trend is the development of large-scale coupled biogeophysical-chemical modules that couple dynamic feedbacks of climate, chemistry, and radiative forcing in the whole framework for air quality forecasting at multiple spatial and temporal scales. It is hypothesized that, in the long term, should DL make significant strides, there will be the potential for integration with numerical models at a deeper level, enhancing and refining them in a mutually beneficial cycle. Although real-time forecasting is a primary application, the successful implementation of cross-sphere coupled models has the potential to significantly enhance scientific discovery. By facilitating enhanced attribution and interpretation of feedback processes, these models would have the capacity to reveal new insights into the mechanisms driving atmospheric dynamics, thereby enhancing our understanding of Earth system interactions. In conclusion, addressing these challenges in concert will pave the way for next-generation DL-based air quality forecasting systems. Such robust and transparent forecasts will be invaluable for policymakers and communities worldwide, enabling proactive measures to protect public health and combat air pollution in the years to come.

Conflict of Interests Hongliang Zhang is an editorial board member of *Frontiers of Environmental Science & Engineering*. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 42377098) and the National Key Research and Development Program of China (No. 2022YFC3701100).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abnar S, Zuidema W H (2020). Quantifying attention flow in transformers. In: Proceedings of the 58th Annual Meeting of the

Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics

Agapiou A (2017). Remote sensing heritage in a petabyte-scale: satellite data and heritage earth engine© applications. International Journal of Digital Earth, 10(1): 85–102

Appel K W, Bash J O, Fahey K M, Foley K M, Gilliam R C, Hogrefe C, Hutzell W T, Kang D W, Mathur R, Murphy B N, et al. (2021). The Community Multiscale Air Quality (CMAQ) model versions 5.3 and 5.3.1: system updates and evaluation. Geoscientific Model Development, 14(5): 2867–2897

Bai K X, Li K, Shao L Q, Li X R, Liu C S, Li Z Q, Ma M L, Han D, Sun Y B, Zheng Z, et al. (2024). LGHAP v2: a global gap-free aerosol optical depth and $PM_{2.5}$ concentration dataset since 2000 derived via big Earth data analytics. Earth System Science Data, 16(5): 2425–2448

Bai L B, Du S H, Zhang X Y, Wang H Y, Liu B, Ouyang S (2022). Domain adaptation for remote sensing image semantic segmentation: an integrated approach of contrastive learning and adversarial learning. IEEE Transactions on Geoscience and Remote Sensing, 60: 5628313

Baklanov A, Brunner D, Carmichael G, Flemming J, Freitas S, Gauss M, Hov Ø, Mathur R, Schlünzen K H, Seigneur C, et al. (2017). Key issues for seamless integrated chemistry–meteorology modeling. Bulletin of the American Meteorological Society, 98(11): 2285–2292

Baklanov A, Schlünzen K, Suppan P, Baldasano J, Brunner D, Aksoyoglu S, Carmichael G, Douros J, Flemming J, Forkel R, et al. (2014). Online coupled regional meteorology chemistry models in Europe: current status and prospects. Atmospheric Chemistry and Physics, 14(1): 317–398

Baró R, Palacios-Peña L, Baklanov A, Balzarini A, Brunner D, Forkel R, Hirtl M, Honzak L, Pérez J L, Pirovano G, et al. (2017). Regional effects of atmospheric aerosols on temperature: an evaluation of an ensemble of online coupled models. Atmospheric Chemistry and Physics, 17(15): 9677–9696

Beucler T, Pritchard M, Rasp S, Ott J, Baldi P, Gentine P (2021). Enforcing analytic constraints in neural networks emulating physical systems. Physical Review Letters, 126(9): 098302

Bi J Z, Knowland K E, Keller C A, Liu Y (2022). Combining machine learning and numerical simulation for high-resolution $PM_{2.5}$ concentration forecast. Environmental Science & Technology, 56(3): 1544–1556

Bi K F, Xie L X, Zhang H H, Chen X, Gu X T, Tian Q (2023). Accurate medium-range global weather forecasting with 3D neural networks. Nature, 619(7970): 533–538

Bodnar C, Bruinsma W P, Lucic A, Stanley M, Allen A, Brandstetter J, Garvan P, Riechert M, Weyn J A, Dong H Y, et al. (2025). A foundation model for the Earth system. Nature, 641(8065): 1180–1187

Bonev B, Kurth T, Hundt C, Pathak J, Baust M, Kashinath K, Anandkumar A (2023). Spherical Fourier neural operators: learning stable dynamics on the sphere. In: Proceedings of the 40th International Conference on Machine Learning. Honolulu: JMLR.org, 117

Bucsela E J, Krotkov N A, Celarier E A, Lamsal L N, Swartz W H, Bhartia P K, Boersma K F, Veefkind J P, Gleason J F, Pickering K E (2013). A new stratospheric and tropospheric NO_2 retrieval algorithm for nadir-viewing satellite instruments: applications to OMI. Atmospheric Measurement Techniques, 6(10): 2607–2626

Buizza R (2019). Introduction to the special issue on "25 Years of Ensemble Forecasting". Quarterly Journal of the Royal Meteorological Society, 145(S1): 1–11

Byun D, Schere K L (2006). Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (CMAQ) modeling system. Applied Mechanics Reviews, 59(2): 51–77

Caldwell P M, Mametjanov A, Tang Q, Van Roekel L P, Golaz J C, Lin W Y, Bader D C, Keen N D, Feng Y, Jacob R, et al. (2019). The DOE E3SM coupled model version 1: description and results at high resolution. Journal of Advances in Modeling Earth Systems, 11(12): 4095–4146

Chen J X, Zhu S Q, Wang P, Zheng Z H, Shi S, Li X Y, Xu C, Yu K X, Chen R J, Kan H D, et al. (2024a). Predicting particulate matter, nitrogen dioxide, and ozone across great britain with high spatiotemporal resolution based on random forest models. Science of the Total Environment, 926: 171831

Chen K, Han T, Ling F H, Gong J C, Bai L, Wang X Y, Luo J J, Fei B, Zhang W L, Chen X, et al. (2025). The operational medium-range deterministic weather forecasting can be extended beyond a 10-day lead time. Communications Earth & Environment, 6(1): 518

Chen K H, Li G B, Li H W, Wang Y Q, Wang W Z, Liu Q Y, Wang H C (2024b). Quantifying uncertainty: air quality forecasting based on dynamic spatial-temporal denoising diffusion probabilistic model. Environmental Research, 249: 118438

Chen L, Zhong X H, Li H, Wu J, Lu B, Chen D L, Xie S P, Wu L B, Chao Q C, Lin C S, et al. (2024c). A machine learning model that outperforms conventional global subseasonal forecast models. Nature Communications, 15(1): 6425

Chen W Y, Lu X C, Yuan D H, Chen Y A, Li Z N, Huang Y Q, Fung T, Sun H C, Fung J C H (2023). Global $PM_{2.5}$ prediction and associated mortality to 2100 under different climate change scenarios. Environmental Science & Technology, 57(27): 10039–10052

Christopher S A, Gupta P (2010). Satellite remote sensing of particulate matter air quality: the cloud-cover problem. Journal of the Air & Waste Management Association, 60(5): 596–602

Cohen A J, Brauer M, Burnett R, Anderson H R, Frostad J, Estep K, Balakrishnan K, Brunekreef B, Dandona L, Dandona R, et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. The Lancet, 389(10082): 1907–1918

Collins W J, Lamarque J F, Schulz M, Boucher O, Eyring V, Hegglin M I, Maycock A, Myhre G, Prather M, Shindell D, et al. (2017). AerChemMIP: quantifying the effects of chemistry and

aerosols in CMIP6. Geoscientific Model Development, 10(2): 585–607

Copernicus Sentinel-5P (processed by ESA) (2020). TROPOMI Level 2 Formaldehyde Total Column products. Version 02. Paris: European Space Agency

Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y J, Schwartz J (2016). Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. Environmental Science & Technology, 50(9): 4712–4721

Domeisen D I V, White C J, Afargan-Gerstman H, Muñoz Á G, Janiga M A, Vitart F, Wulff C O, Antoine S, Ardilouze C, Batté L, et al. (2022). Advances in the subseasonal prediction of extreme events: relevant case studies across the globe. Bulletin of the American Meteorological Society, 103(6): E1473–E1501

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. (2021). An image is worth 16 × 16 words: transformers for image recognition at scale. In: Proceedings of the 9th International Conference on Learning Representations. Virtual Only Conference, Vienna, Austria

Ehrendorfer M (1997). Predicting the uncertainty of numerical weather forecasts: a review. Meteorologische Zeitschrift, 6(4): 147–183

Fan X, Wang Z H, Lin Y T, Zhang Y, Xiang Y, Li H (2025). MVAR: MultiVariate Auto Regressive Air Pollutants Forecasting Model. Shanghai: Shanghai Academy of Artificial Intelligence for Science, arxiv.org/abs/2507.12023

Franchi G, Laurent O, Leguéry M, Bursuc A, Pilzer A, Yao A (2024). Make me a BNN: a simple strategy for estimating Bayesian uncertainty from pre-trained models. In: Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 12194–12204

Gal Y, Ghahramani Z (2016). Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning. New York: JMLR.org, 1050–1059

Geng G N, Xiao Q Y, Liu S G, Liu X D, Cheng J, Zheng Y X, Xue T, Tong D, Zheng B, Peng Y R, et al. (2021). Tracking air pollution in China: near real-time $PM_{2.5}$ retrievals from multisource data fusion. Environmental Science & Technology, 55(17): 12106–12115

Griffiths P T, Wilcox L J, Allen R J, Naik V, O'Connor F M, Prather M J, Archibald A T, Brown F, Deushi M, Collins W, et al. (2024). The role of AerChemMIP in advancing climate and air quality research. Atmospheric Chemistry and Physics, 25, 8289–8328

Guastavino S, Piana M, Tizzi M, Cassola F, Iengo A, Sacchetti D, Solazzo E, Benvenuto F (2022). Prediction of severe thunderstorm events with ensemble deep learning and radar data. Scientific Reports, 12(1): 20049

Hammer M S, Van Donkelaar A, Bindle L, Sayer A M, Lee J, Hsu N C, Levy R C, Sawyer V, Garay M J, Kalashnikova O V, et al. (2023). Assessment of the impact of discontinuity in satellite instruments and retrievals on global $PM_{2.5}$ estimates. Remote Sensing of Environment, 294: 113624

He K M, Chen X L, Xie S N, Li Y H, Dollár P, Girshick R (2022a). Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 15979–15988

He T L, Jones D B A, Miyazaki K, Bowman K W, Jiang Z, Chen X K, Li R, Zhang Y X, Li K N (2022b). Inverse modelling of Chinese NO_x emissions using deep learning: integrating *in situ* observations with a satellite-based chemical reanalysis. Atmospheric Chemistry and Physics, 22(21): 14059–14074

Hettige K H, Ji J H, Xiang S L, Long C, Cong G, Wang J Y (2024). AirPhyNet: harnessing physics-guided neural networks for air quality prediction. In: Proceedings of the 12th International Conference on Learning Representations. Vienna: ICLR

Hirsch E, Koren I (2021). Record-breaking aerosol levels explained by smoke injection into the stratosphere. Science, 371(6535): 1269–1274

Ho J, Jain A, Abbeel P (2020). Denoising diffusion probabilistic models. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 574

Hong C P, Zhang Q, Zhang Y, Davis S J, Tong D, Zheng Y X, Liu Z, Guan D B, He K B, Schellnhuber H J (2019). Impacts of climate change on future air quality and human health in China. Proceedings of the National Academy of Sciences of the United States of America, 116(35): 17193–17200

Hu M Y, Lu X C, Chen Y A, Li Z N, Wang Y Y, Fung J C H (2025). AirQFormer: improving regional air quality forecast with a hybrid deep learning model. Sustainable Cities and Society, 119: 106113

Huang C H, Hu J L, Xue T, Xu H, Wang M (2021a). High-resolution spatiotemporal modeling for ambient $PM_{2.5}$ exposure assessment in China from 2013 to 2019. Environmental Science & Technology, 55(3): 2152–2162

Huang C Y, Hu T T, Duan Y S, Li Q Y, Chen N, Wang Q, Zhou M G, Rao P H (2022). Effect of urban morphology on air pollution distribution in high-density urban blocks based on mobile monitoring and machine learning. Building and Environment, 219: 109173

Huang L, Liu S, Yang Z Y, Xing J, Zhang J, Bian J, Li S W, Sahu S K, Wang S X, Liu T Y (2021b). Exploring deep learning for air pollutant emission estimation. Geoscientific Model Development, 14(7): 4641–4654

Jairi I, Ben-Othman S, Canivet L, Zgaya-Biau H (2024). Enhancing air pollution prediction: a neural transfer learning approach across different air pollutants. Environmental Technology & Innovation, 36: 103793

Jiang F, Zheng Z H, Coe H, Healy R M, Poulain L, Gros V, Zhang H, Li W J, Liu D T, West M, et al. (2025). Integrating simulations and observations: a foundation model for estimating the aerosol mixing state index. ACS ES&T Air, 2(5): 877–890

Karras T, Aittala M, Laine S, Aila T (2022). Elucidating the design

space of diffusion-based generative models. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 1926

Kelp M M, Jacob D J, Kutz J N, Marshall J D, Tessum C W (2020). Toward stable, general machine-learned models of the atmospheric chemical system. Journal of Geophysical Research: Atmospheres, 125(23): e2020JD032759

Kelp M M, Jacob D J, Lin H P, Sulprizio M P (2022). An online-learned neural network chemical solver for stable long-term global simulations of atmospheric chemistry. Journal of Advances in Modeling Earth Systems, 14(6): e2021MS002926

Kochkov D, Yuval J, Langmore I, Norgaard P, Smith J, Mooers G, Klöwer M, Lottes J, Rasp S, Düben P, et al. (2024). Neural general circulation models for weather and climate. Nature, 632(8027): 1060–1066

Lam R, Sanchez-Gonzalez A, Willson M, Wirnsberger P, Fortunato M, Alet F, Ravuri S, Ewalds T, Eaton-Rosen Z, Hu W H, et al. (2023). Learning skillful medium-range global weather forecasting. Science, 382(6677): 1416–1421

Lamarque J F, Emmons L K, Hess P G, Kinnison D E, Tilmes S, Vitt F, Heald C L, Holland E A, Lauritzen P H, Neu J, et al. (2012). CAM-chem: description and evaluation of interactive atmospheric chemistry in the community earth system model. Geoscientific Model Development, 5(2): 369–411

Larkin A, Anenberg S, Goldberg D L, Mohegh A, Brauer M, Hystad P (2023). A global spatial-temporal land use regression model for nitrogen dioxide air pollution. Frontiers in Environmental Science, 11: 1125979

Li J W, Han Z W, Surapipith V, Fan W X, Thongboonchoo N, Wu J, Li J, Tao J, Wu Y F, Macatangay R, et al. (2022a). Direct and indirect effects and feedbacks of biomass burning aerosols over mainland southeast Asia and south China in springtime. Science of the Total Environment, 842: 156949

Li J Y, Cai Y X, Li Q, Kou M Y, Zhang T X (2024a). A review of remote sensing image segmentation by deep learning methods. International Journal of Digital Earth, 17(1): 2328827

Li M Y, Yang Q Q, Yuan Q Q, Zhu L Y (2022b). Estimation of high spatial resolution ground-level ozone concentrations based on Landsat 8 TIR bands with deep forest model. Chemosphere, 301: 134817

Li S W, Xing J (2025). Enhancing 72-hour air quality forecasting with an observation-driven deep learning chemistry transport model. Environment International, 202: 109689

Li Z Y, Han W, Zhang Y, Fu Q F, Li J X, Qin L Z, Dong R Y, Sun H, Deng Y, Yang L J (2024b). Learning spatiotemporal dynamics with a pretrained generative model. Nature Machine Intelligence, 6(12): 1566–1579

Libonati R, Geirinhas J L, Silva P S, Russo A, Rodrigues J A, Belém L B C, Nogueira J, Roque F O, DaCamara C C, Nunes A M B, et al. (2022). Assessing the role of compound drought and heatwave events on unprecedented 2020 wildfires in the Pantanal. Environmental Research Letters, 17(1): 015005

Lin Y (2024). Progressive neural network for multi-horizon time series forecasting. Information Sciences, 661: 120112

Liu H, Xu Y N, Chen C (2019). Improved pollution forecasting hybrid algorithms based on the ensemble method. Applied Mathematical Modelling, 73: 473–486

Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S, Guo B N (2021a). Swin transformer: hierarchical vision transformer usingshifted windows. In: Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 9992–1000

Liu N, Xiong A Y, Zhang Q, Liu Y J, Zhan Y J, Liu Y M (2021b). Development of basic dataset of severe convective weather for artificial intelligence training. Journal of Applied Meteorological Science, 32(5): 530–541

Liu W Q, Xing C Z (2024). Needs and challenges of optical atmospheric monitoring on the background of carbon neutrality in China. Frontiers of Environmental Science & Engineering, 18(6): 73

Liu X, Zhu Y J, Xue L, Desai A R, Wang H K (2022). Cluster-enhanced ensemble learning for mapping global monthly surface ozone from 2003 to 2019. Geophysical Research Letters, 49(11): e2022GL097947

Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S, Guo B N (2021). Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 9992–10002

Lu M Y, Lao T F, Yu M Z, Zhang Y D, Zheng J Q, Li Y C (2021). $PM_{2.5}$ concentration forecasting over the central area of the Yangtze River delta based on deep learning considering the spatial diffusion process. Remote Sensing, 13(23): 4834

Lu X, Zhang S J, Xing J, Wang Y J, Chen W H, Ding D, Wu Y, Wang S X, Duan L, Hao J M (2020). Progress of air pollution control in China and its challenges and opportunities in the ecological civilization era. Engineering, 6(12): 1423–1431

Lundberg S M, Lee S I (2017). A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 4768–4777

Lyapustin A, Wang Y (2022). MODIS/Terra+Aqua Land Aerosol Optical Depth Daily L2G Global 1km SIN Grid. Washington, DC: NASA EOSDIS Land Processes DAAC

Lyu B, Huang R, Wang X L, Wang W G, Hu Y T (2024). FastCTM (v1.0): atmospheric chemical transport modelling with a principle-informed neural network for air quality simulations. Geoscientific Model Development Discussions, doi: 10.5194/gmd-2024-1982024

Ma J, Li Z, Cheng J C P, Ding Y X, Lin C Q, Xu Z R (2020). Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. Science of the Total Environment, 705: 135771

Mariotti A, Ruti P M, Rixen M (2018). Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. npj Climate and Atmospheric Science, 1(1): 4

Mechoso C R (2020). Interacting Climates of Ocean Basins: Observations, Mechanisms, Predictability, and Impacts. Cambridge: Cambridge University Press

Miyazaki K, Eskes H J, Sudo K (2012). Global NO_x emission estimates derived from an assimilation of OMI tropospheric NO_2 columns. Atmospheric Chemistry and Physics, 12(5): 2263–2288

Mo X Y, Li H, Zhang L (2022). Design a regional and multistep air quality forecast model based on deep learning and domain knowledge. Frontiers in Earth Science, 10: 995843

Molnar C (2025). Interpretable Machine Learning: a Guide for Making Black Box Models Explainable. 3rd ed. ISBN: 978-3-911578-03-5

Montavon G, Samek W, Müller K R (2018). Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 73: 1–15

Price I, Sanchez-Gonzalez A, Alet F, Andersson T R, El-Kadi A, Masters D, Ewalds T, Stott J, Mohamed S, Battaglia P, et al. (2025). Probabilistic weather forecasting with machine learning. Nature, 637(8044): 84–90

Ran N, Xiao P, Wang Y, Shi W, Lin J X, Meng Q, Allmendinger R (2025). HR-Extreme: a high-resolution dataset for extreme weather forecasting. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: ICLR

Ravi R, Kumari N S, Geethika P S S, Rao K V, Rao M S (2024). Air pollution forecasting using deep learning algorithms: a review. In: Lin F M, Patel A, Kesswani N, Sambana B, eds. Accelerating Discoveries in Data Science and Artificial Intelligence I. Cham: Springer, 511–517

Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743): 195–204

Ren X, Mi Z Y, Cai T, Nolte C G, Georgopoulos P G (2022). Flexible Bayesian ensemble machine learning framework for predicting local ozone concentrations. Environmental Science & Technology, 56(7): 3871–3883

Requia W J, Di Q, Silvern R, Kelly J T, Koutrakis P, Mickley L J, Sulprizio M P, Amini H, Shi L H, Schwartz J (2020). An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. Environmental Science & Technology, 54(18): 11037–11047

Schreck J S, Gagne D J, Becker C, Chapman W E, Elmore K, Fan D, Gantos G, Kim E, Kimpara D, Martin T, et al. (2024). Evidential deep learning: enhancing predictive uncertainty estimation for Earth system science applications. Artificial Intelligence for the Earth Systems, 3(4): 230093

Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2020). Grad-CAM: visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision, 128(2): 336–359

Sha Y K, Sobash R A, Gagne D J (2024). Generative ensemble deep learning severe weather prediction from a deterministic convection-allowing model. Artificial Intelligence for the Earth Systems, 3(2): e230094

Shen L, Mickley L J (2017). Seasonal prediction of US summertime ozone using statistical analysis of large scale climate patterns. Proceedings of the National Academy of Sciences of the United States of America, 114(10): 2491–2496

Shen S Y, Li C, van Donkelaar A, Jacobs N, Wang C G, Martin R V (2024). Enhancing global estimation of fine particulate matter concentrations by including geophysical *a priori* information in deep learning. ACS ES&T Air, 1(5): 332–345

Song Z H, Chen B, Zhang P, Guan X D, Wang X, Ge J M, Hu X Q, Zhang X Y, Wang Y X (2022). High temporal and spatial resolution $PM_{2.5}$ dataset acquisition and pollution assessment based on FY-4A TOAR data and deep forest model in China. Atmospheric Research, 274: 106199

Tao C L, Jia M, Wang G Q, Zhang Y Q, Zhang Q Z, Wang X F, Wang Q, Wang W X (2024a). Time-sensitive prediction of NO_2 concentration in China using an ensemble machine learning model from multi-source data. Journal of Environmental Sciences, 137: 30–40

Tao C L, Peng Y B, Zhang Q Z, Zhang Y Q, Gong B, Wang Q, Wang W X (2024b). Diagnosing ozone-NO_x-VOC-aerosol sensitivity and uncovering causes of urban-nonurban discrepancies in Shandong, China, using transformer-based estimations. Atmospheric Chemistry and Physics, 24(7): 4177–4192

Thongthammachart T, Araki S, Shimadera H, Matsuo T, Kondo A (2022). Incorporating Light Gradient Boosting Machine to land use regression model for estimating NO_2 and $PM_{2.5}$ levels in Kansai region, Japan. Environmental Modelling & Software, 155: 105447

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł, Polosukhin I (2017). Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 6000–6010

Wang C G, Pritchard M S, Brenowitz N, Cohen Y, Bonev B, Kurth T, Durran D, Pathak J (2024). Coupled Ocean-Atmosphere Dynamics in a Machine Learning Earth System Model. Princeton: Princeton University

Wang Y, Puķīte J, Wagner T, Donner S, Beirle S, Hilboll A, Vrekoussis M, Richter A, Apituley A, Piters A, et al. (2018). Vertical profiles of tropospheric ozone from MAX-DOAS measurements during the CINDI-2 campaign. Part 1-development of a new retrieval algorithm. Journal of Geophysical Research: Atmospheres, 123(18): 10637–10670

Wang Z X, Li J, Wu L, Zhu M M, Zhang Y J, Ye Z L, Wang Z F (2022). Deep learning-based gas-phase chemical kinetics kernel emulator: application in a global air quality simulation case. Frontiers in Environmental Science, 10: 955980

Watanachaturaporn P, Arora M, Varshney P (2006). Sub-Pixel Land Cover Classification Using Support Vector Machines. In: ASPRS 2006 Annual Conference. May 1–5, Reno, Nevada

Wei J, Li Z Q, Lyapustin A, Wang J, Dubovik O, Schwartz J, Sun L, Li C, Liu S, Zhu T (2023). First close insight into global daily

gapless 1 km $PM_{2.5}$ pollution, variability, and health impact. Nature Communications, 14(1): 8349

Wei J, Liu S, Li Z Q, Liu C, Qin K, Liu X, Pinker R T, Dickerson R R, Lin J T, Boersma K F, et al. (2022). Ground-level NO_2 surveillance from space across China for high resolution using interpretable spatiotemporally weighted artificial intelligence. Environmental Science & Technology, 56(14): 9988–9998

Whittaker T, Di Luca A (2025). Pushing the Limits of Extreme Weather: Constructing Extreme Heatwave Storylines with Differentiable Climate Models. Montréal: Université du Québec à Montréal, arXiv:2506.10660

Xie Y Y, Lin M Y, Decharme B, Delire C, Horowitz L W, Lawrence D M, Li F, Séférian R (2022). Tripling of western US particulate pollution from wildfires in a warming climate. Proceedings of the National Academy of Sciences of the United States of America, 119(14): e2111372119

Xing J, Li S, Zheng S, Liu C, Wang X, Huang L, Song G, He Y, Wang S, Sahu S K, et al. (2022). Rapid inference of nitrogen oxide emissions based on a top-down method with a physically informed variational autoencoder. Environmental Science & Technology, 56(14): 9903–9914

Xing J, Zheng S X, Ding D, Kelly J T, Wang S X, Li S W, Qin T, Ma M Y, Dong Z X, Jang C, et al. (2020). Deep learning for prediction of the air quality response to emission changes. Environmental Science & Technology, 54(14): 8589–8600

Xiong Y, Yang Q Q, Gao Y, Li K, Yang Y, Lin G X, Lu X, Wang Z L, Zhang H L, Gao M (2025). Modeling the formation of aerosols and their interactions with weather and climate: critical review and future perspectives. Frontiers of Environmental Science & Engineering, 19(11): 143

Xu J Z, Zhang H R, Cheng Z, Liu J Y, Xu Y Y, Wang Y C (2022). Approximating three-dimensional (3-D) transport of atmospheric pollutants via deep learning. Earth and Space Science, 9(7): e2022EA002338

Zhan J L, Liu Y C, Ma W, Zhang X, Wang X Z, Bi F, Zhang Y J, Wu Z H, Li H (2022). Ozone formation sensitivity study using machine learning coupled with the reactivity of volatile organic compound species. Atmospheric Measurement Techniques, 15(5): 1511–1520

Zhang A X, Fu T M, Feng X, Guo J F, Liu C F, Chen J K, Mo J J, Zhang X, Wang X L, Wu W L, et al. (2023). Deep learning-based ensemble forecasts and predictability assessments for surface ozone pollution. Geophysical Research Letters, 50(8): e2022GL102611

Zhang B R (2025). Comparative investigation of machine learning and deep learning approaches for air quality prediction. ITM Web of Conferences, 73: 02002

Zhang C X, Niu X H, Wu H Y, Ding Z P, Chan K L, Kim J, Wagner T, Liu C (2025). Unleashing the potential of geostationary satellite observations in air quality forecasting through artificial intelligence techniques. Atmospheric Chemistry and Physics, 25(2): 759–770

Zhang Z, Zhang S (2023). Modeling air quality $PM_{2.5}$ forecasting using deep sparse attention-based transformer networks. International Journal of Environmental Science and Technology, 20(12): 13535–13550

Zhao N, Zhang Y Q, Xue L K (2025). Nonlinear relationship between air pollution and precursor emissions in Qingdao, eastern China. Frontiers of Environmental Science & Engineering, 19(1): 9

Zheng Z H, Fiore A M, Westervelt D M, Milly G P, Goldsmith J, Karambelas A, Curci G, Randles C A, Paiva A R, Wang C, et al. (2023). Automated machine learning to evaluate the information content of tropospheric trace gas columns for fine particle estimates over India: a modeling testbed. Journal of Advances in Modeling Earth Systems, 15(3): e2022MS003099

Zhong Q R, Schutgens N, Veraverbeke S, van der Werf G R (2024). Increasing aerosol emissions from boreal biomass burning exacerbate arctic warming. Nature Climate Change, 14(12): 1275–1281

Zhou Y L, Chang F J, Chang L C, Kao I F, Wang Y S (2019). Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. Journal of Cleaner Production, 209: 134–145

Zhu Q Y, Bi J Z, Liu X, Li S S, Wang W H, Zhao Y, Liu Y (2022). Satellite-based long-term spatiotemporal patterns of surface ozone concentrations in China: 2005-2019. Environmental Health Perspectives, 130(2): 027004

Zhu S Q, Ma J L, Wang S Y, Sun S D, Wang P, Zhang H L (2023). Shifts of formation regimes and increases of atmospheric oxidation led to ozone increase in North China Plain and Yangtze River Delta from 2016 to 2019. Journal of Geophysical Research: Atmospheres, 128(13): e2022JD038373