Geosci. Model Dev., 18, 8777–8800, 2025
https://doi.org/10.5194/gmd-18-8777-2025

Geoscientific Model Development
Open Access
EGU

Applications of Machine Learning and Artificial Intelligence in Tropospheric Ozone Research

Sebastian H. M. Hickman[1,★], Makoto M. Kelp[2,★], Paul T. Griffiths[3,a,★], Kelsey Doerksen[4,★], Kazuyuki Miyazaki[5,★], Elyse A. Pennington[5,★], Gerbrand Koren[6,★], Fernando Iglesias-Suarez[7,★], Martin G. Schultz[8,16,★], Kai-Lan Chang[9,10], Owen R. Cooper[9], Alex Archibald[3], Roberto Sommariva[11,12], David Carlson[13], Hantao Wang[14], J. Jason West[14], and Zhenze Liu[15]

[1]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, Cambridge, UK
[2]Doerr School of Sustainability, Stanford University, Stanford, CA, USA
[3]National Centre for Atmospheric Science, University of Cambridge, Cambridge, UK
[4]Department of Computer Science, University of Oxford, Oxford, UK
[5]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[6]Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, the Netherlands
[7]Predictia Intelligent Data Solutions S.L., Santander, Spain
[8]Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany
[9]NOAA Chemical Sciences Laboratory, Boulder, CO, USA
[10]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA
[11]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK
[12]School of Chemistry, University of Leicester, Leicester, UK
[13]Civil and Environmental Engineering, Duke University, Durham, NC, USA
[14]Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill, NC, USA
[15]Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China
[16]Department of Mathematics and Computer Science, University of Cologne, Cologne, Germany
[a]now at: School of Chemistry, Bristol University, Bristol, UK
★These authors contributed equally to this work.

Correspondence: Paul T. Griffiths (paul.griffiths@ncas.ac.uk)

Received: 29 November 2024 – Discussion started: 6 January 2025
Revised: 29 July 2025 – Accepted: 6 October 2025 – Published: 20 November 2025

Abstract. Machine learning (ML) is transforming atmospheric chemistry, offering powerful tools to address challenges in tropospheric ozone research, a critical area for climate resilience and public health. As in adjacent fields, ML approaches complement existing research by learning patterns from ever-increasing volumes of atmospheric and environmental data relevant to ozone. We highlight the rapid progress made in the field since Phase 1 of the Tropospheric Ozone Assessment Report (TOAR), focussing particularly on the most active areas of research, namely short-term ozone forecasting, emulation of atmospheric chemistry and the use of remote sensing for ozone estimation. This review provides a comprehensive synthesis of recent advancements, highlights critical challenges, and proposes actionable pathways to develop ML in ozone research. Further advances hinge on addressing domain-specific issues such as the dependence of ozone concentrations on several poorly observed precursor species, as well as making progress on generic ML challenges such as the definition of suitable benchmarks and developing robust, explainable models. Reaping the full potential of ML for ozone research and operational applications will require close collaborations across atmospheric chemistry, ML and computational science and vigilant pursuit of the rapid developments in adjacent fields.

Published by Copernicus Publications on behalf of the European Geosciences Union.

1 Introduction

The ML4O3 working group was established as part of the second phase of the IGAC Tropospheric Ozone Assessment Report (TOAR). The group focuses on the application of machine learning (ML) concepts and methods, promoting dialogue between researchers in machine learning and tropospheric ozone communities. The motivation of this group is to allow the atmospheric chemistry community to capitalize on the potential of ML and AI techniques that has recently been demonstrated for weather and climate applications. The ML tasks that were addressed by the group included identifying complex patterns, interpolating missing values, detecting errors or anomalies, and identifying air pollution regimes. The working group aimed to contribute to both fundamental scientific understanding of the processes controlling ozone, and to improved air quality monitoring and forecasting.

Tropospheric ozone is a harmful atmospheric pollutant and an important greenhouse gas, contributing to both environmental and public health issues. Long-term exposure to elevated ozone levels is linked to hundreds of thousands of premature deaths globally each year (Malashock et al., 2022; Malley et al., 2017; Health Effects Institute, 2024). Short-term exposure can cause serious negative health impacts (Bell et al., 2014) including reduced lung function, particularly in individuals with pre-existing medical conditions (EPA, 2020). Beyond its health impacts, tropospheric ozone significantly damages vegetation in natural ecosystems and agricultural fields (Mills et al., 2018) and can act as a climate forcer in the upper troposphere. In addition, ozone plays a critical role in tropospheric chemistry, both as a source of oxidants and as a primary oxidant itself (Monks et al., 2015).

Ozone is challenging to simulate accurately (Young et al., 2018), also for ML models, because it is not directly emitted into the troposphere but is photochemically produced in the presence of sunlight by reactions involving its precursor gases: carbon monoxide (CO), methane (CH_4), volatile organic compounds (VOCs), and nitrogen oxides (NO_x, $NO+NO_2$). In addition, ozone is transported from the stratosphere into the troposphere. The removal of tropospheric ozone is controlled by chemical loss and deposition to the surface (Archibald et al., 2020). The lifetime of ozone in the troposphere ranges from days to weeks, depending on local chemical and meteorological conditions (Lelieveld and Dentener, 2000; Monks et al., 2015). This variability allows ozone and its precursors to be transported over long distances from their sources (Fiore et al., 2009).

The complex coupling of these chemical and physical processes controls the local concentrations of ozone across different spatial and temporal scales, as detailed in Fig. 1. Traditionally, concentrations of ozone and other chemical species are calculated using numerical models of the atmosphere that represent these processes across a wide range of spatial scales, from high-resolution urban models (meter-scale) to global chemistry-climate models with resolutions ranging from tens to hundreds of kilometers (e.g. Morgenstern et al., 2017).

Despite the success of ozone simulations in air quality and climate research, large uncertainties still exist in global model estimates of tropospheric ozone and its trends, although ozone is the longest- and most-measured trace gas in the observational record. Observations from ground stations, ozonesondes, and satellites indicate that tropospheric ozone has generally increased in recent decades (Ziemke et al., 2019; Young et al., 2018; Gulev et al., 2021). While global atmospheric chemistry models agree that the global tropospheric ozone burden has increased from pre-industrial times to the present day, they vary regarding the spatial distribution and magnitude of the increase (Skeie et al., 2020; Christiansen et al., 2022; Fiore et al., 2022). Potential sources driving this model bias include uncertainties in tropical emissions (Zhang et al., 2021), nonlinear NO_x-VOC chemistry (Shah et al., 2023), stratosphere-troposphere exchange (Neu et al., 2014), boundary layer mixing (Lu et al., 2019), missing chemical mechanisms such as halogen chemistry (Wang et al., 2015), and deposition (Clifton et al., 2020).

The variation of ozone at various scales is shown in Fig. 2. The figure shows the diurnal and annual cycles of ozone at four sites from the TOAR database: Mauna Loa Observatory, a Pacific mountain station, based in Hawaii, USA; Minamitorishima, a Pacific island station in Japan; a regional continental background site, Borken, Germany, and an urban, roadside site, Marylebone Road, London, UK. There is little consistency between the diurnal cycles at the various sites: the remote Pacific site sees little diurnal variation in ozone, but a strong seasonal cycle, with levels reaching a minimum in the summer. In contrast, the continental, rural background site in Germany has a strong diurnal cycle, peaking in the late afternoon, and a strong seasonal cycle with a summertime maximum. The observed long-term trends in ozone, although weak, also vary between the sites, with both modest increases (London Marylebone Road) and decreases (Minamitorishima). The lower panel shows variation in ozone trends across the globe, ranging between -3 and $3\,\mathrm{ppbv\,yr^{-1}}$, across remote sites. These differences across ozone monitoring sites result from the complex interactions between precursor emissions, transport and chemical processes, meteorological drivers, and surface characteristics. Capturing the diversity of daily and seasonal cycles as well as trends is a key requisite of any ozone model.

Machine learning (ML) approaches, which can learn and reproduce nonlinear characteristics of a system from data (Hornik et al., 1989), may provide a valuable complement to physical models. As the quantity and quality of observational data on ozone (Schultz et al., 2017) and on the broader Earth system (Agapiou, 2017; Reichstein et al., 2019) continue to grow, ML is becoming an increasingly viable tool for advancing ozone research.

Figure 3 highlights progress in the field of ML as applied to weather and climate science, which has been rapid since

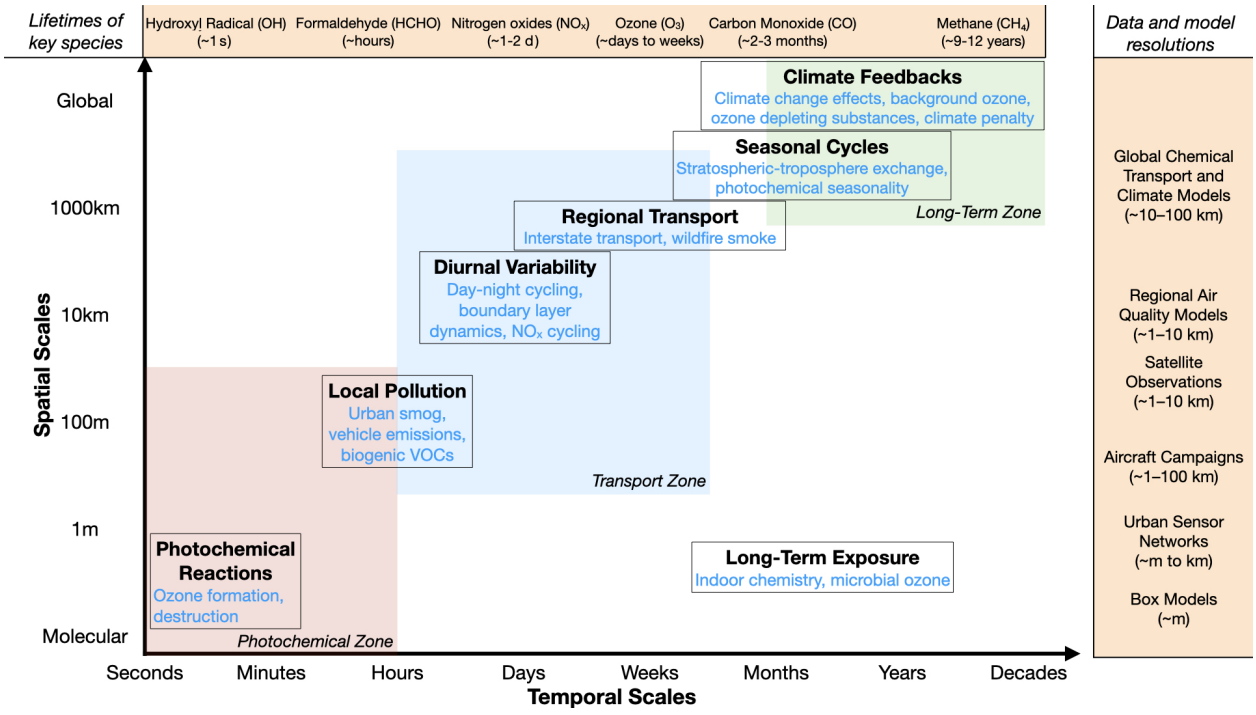

Figure 1. Spatial and temporal scales of tropospheric ozone chemistry processes. The x-axis shows timescales, from rapid photochemical reactions to long-term climate feedbacks, and the y-axis shows spatial scales, from local pollution to global atmospheric transport. Species lifetimes and relevant data sources and models are displayed to illustrate the range and scales of phenomena and methods used to study ozone chemistry.

the publication of the first phase of the TOAR assessment. In their review of the state of the field of weather forecasting, Bauer et al. (2015) note many areas of progress for the field, including model throughput, the process-level detail of then current models, and the use of data assimilation techniques to improve the fidelity of the model's initial state. The impact of ML methods was not anticipated. Rasp et al. (2018) demonstrated the potential for deep learning techniques to augment existing models in providing an alternative, complementary and physically consistent description of sub-grid scale processes, such as cloud microphysics. The coupling of a fast, accurate, data-driven module, trained on finer scale simulations, to a larger scale host climate model exemplified one of the potential ways that ML approaches can contribute to the improvement of climate and weather models. Subsequent studies have shown in various ways the advantages of ML over traditional numerical models, particularly in terms of computational efficiency and in the ability to learn from large datasets, as demonstrated by the success of data-driven nowcasting and weather forecasting models (Bi et al., 2023; Lam et al., 2023; Price et al., 2024).

Observational data, when integrated with model simulations through data assimilation techniques, have improved the understanding of emissions and atmospheric chemistry by reducing uncertainties (Miyazaki et al., 2020). ML can complement these efforts by combining observational data with model outputs, emulating model components, or enabling computationally cheaper simulations, thereby efficiently diagnosing sources of error in global atmospheric models and improving tropospheric ozone estimates. However, ML also has limitations, such as challenges in generalization, validation, and interpretability. Addressing these issues may be particularly relevant for the ozone modeling community where both predictive accuracy and physical understanding are valued.

In this Perspective, we provide an overview of the state of ML in tropospheric ozone research, review previous applications of ML to various problems related to ozone, and discuss persistent challenges and emerging opportunities. We highlight three areas where ML for ozone has been most widely applied: forecasts based on ground-based observations are reviewed in Sect. 2, methods for complementing or replacing parameterizations in numerical models of atmospheric chemistry and transport are discussed in Sect. 3, and ML models that use satellite data or combined data products are presented in Sect. 4. Section 5 highlights and further details these cross-cutting issues and limitations with the application of ML to ozone studies, while Sect. 6 describes future directions for the field, highlighting emerging approaches that seek to address the cross-cutting challenges.

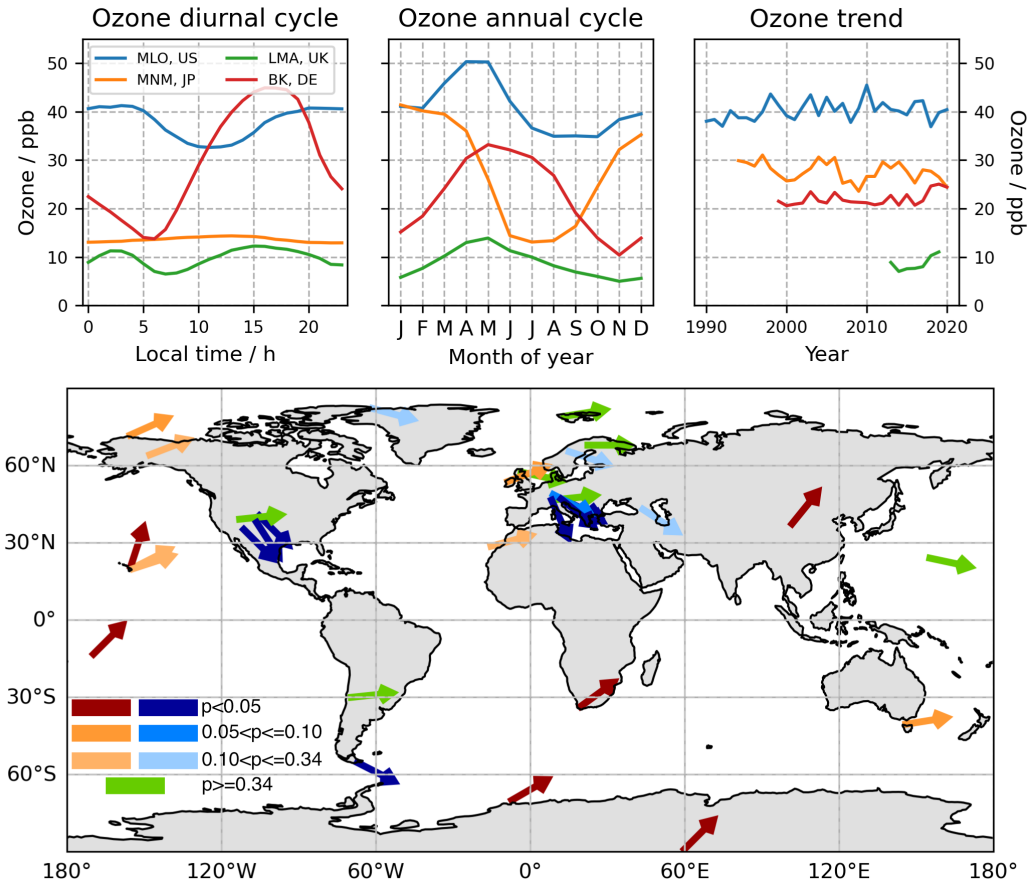

Figure 2. Upper panel: data from the TOAR ozone database for four sites in the northern hemisphere, showing diurnal and seasonal cycles in ozone, and the long-term ozone trend. MLO, US: Mauna Loa Observatory, US; MNM, JP: Minamitorishima, Japan; LMA, UK: Marylebone Road, London, UK; BK, DE: Borken, Germany. Lower panel: long-term ozone trends based on monthly anomalies at remote surface sites. Red and blue indicate positive or negative trends respectively, with different shades giving the statistical significance of the trend at each site. Data from Cooper et al. (2020) and replotted here.

2 Applications of ML to in-situ ozone observations: short-term ground level ozone forecasting

2.1 Background

The short-term forecasting of air pollutants including ozone, i.e. predictions of expected concentrations over 1–4 d, is relevant for public health and scientific questions (Buonocore et al., 2021; Hahm and Yoon, 2021; Alari et al., 2021; Saberian et al., 2017). State-of-the-art air quality forecasts, typically on the timescales of hours to days and up to a few days ahead, are based on the output of numerical chemical transport models (CTMs) (Marécal et al., 2015). These models may be run at higher spatial resolution for the area of interest (Savage et al., 2013), in order to better represent processes controlling air pollution at the local level, and may be post-processed to more accurately represent observations (Casciaro et al., 2022). As with other air pollutants, notably $PM_{2.5}$ (Feng et al., 2015), ML is increasingly being directly applied to the task of short-term, ground-level ozone forecasting, and to bias-correct existing air quality forecasting systems with considerable success. The availability of large and growing observational datasets has facilitated these advances (Schultz et al., 2017). However, forecasting ozone concentrations as time series with ML comes with significant challenges: forecasting ozone is a spatiotemporal problem, and ozone is controlled by processes of varying spatial and temporal scales as shown in Fig. 1.

Many short-term forecasting studies using ML have focused on forecasting only at selected observational stations, using observed ozone and additional chemical species, and meteorological variables where they are available from individual stations or external datasets (Comrie, 1997; Cobourn et al., 2000; Kolehmainen et al., 2001; Eslami et al., 2020; Sayeed et al., 2021; Leufen et al., 2023; Hickman et al., 2023). Furthermore, since it is difficult to downscale a relatively coarse CTM at specific locations, using time series data from a particular station is an attractive way to make predictions at particular locations. However, approaches of this kind do not necessarily provide ozone forecasts across all lo-

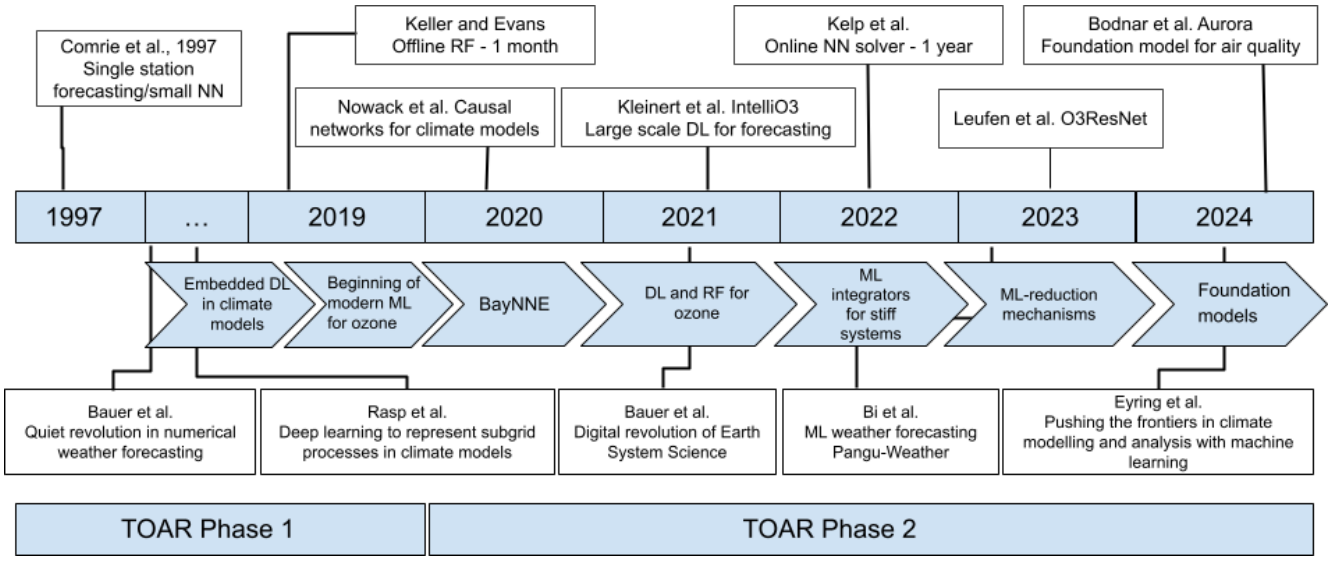

Figure 3. Timeline of a selection of studies using ML in ozone research (top), aligned with a selection of papers using ML in wider weather and climate modeling research (bottom). In both wider Earth system modeling research, and in ozone research there has been rapid progress over the last five years, as noted by landmark review papers highlighted in the Figure. The acronyms used are as follows. NN: neural network; RF: random forest; DL: deep learning; ML: machine learning.

cations that may be of interest, as a gridded model product model might.

2.2 Progress and State of the Science

As with other fields, the advances in ML-based ozone forecasting have been pushed by developments on two axes – first, increasing quantities of data and second, larger models with more appropriate inductive biases. The field has a long history (see Fig. 3), with studies being published even during the most recent artificial intelligence (AI) "winter", beginning with a feed-forward neural network (NN) in 1996 (Yi and Prybutok, 1996). Comrie (1997) illustrated that a NN could be used to forecast ozone at eight stations in the USA. This was followed by further feed-forward NN approaches, often with datasets drawn from a single location or city (Cobourn et al., 2000; Kolehmainen et al., 2001). Neural methods were typically evaluated in comparison with (autoregressive) regression models, often finding that NNs were better able to forecast ozone concentrations and extrema on test data (Nunnari et al., 1998; Schlink et al., 2003; Chaloulakou et al., 2003), although the improvement was often only marginal. Alongside the successes of feed-forward NN architectures, other work drew attention to methods seen to be more interpretable, such as fuzzy logic systems and regression trees (Gardner and Dorling, 2000; Heo and Kim, 2004). Further work leveraged methodological advances in ML architectures designed for temporal data, including the use of recurrent neural networks (RNNs) and convolutional neural networks (CNNs) to account for lagged relationships in the time series data (Eslami et al., 2020; Sayeed et al., 2021; Kleinert et al., 2021). Recent work has combined architectures to model the relationships that control ozone, including combining components such as transformers and CNNs to account for the temporal and spatial information relevant to forecasting (Chen et al., 2022; Cheng et al., 2022; Han et al., 2023). However, datasets have typically been limited to single countries or cities, due to the lack of a combined database of station measurements. The introduction of the TOAR surface database (Schultz et al., 2017) and the TOAR-II database have facilitated recent studies on data drawn from multiple countries (Leufen et al., 2023; Hickman et al., 2023). The importance of the curation of large datasets for scientific progress in ML is highlighted in the Outlook section.

Increasingly, more complex architectures are being used to enhance the accuracy of ozone forecasts, and more data are being included as input to the models. The inputs that are relevant to the physical drivers of ozone concentrations, such as past observations of ozone and covariates, and nearby covariates, reflect processes that control ozone observations, and feasibly contribute to improved ozone forecasting and infilling. Recently, methods on the scale of the ML architectures and data used for weather forecasting (Bi et al., 2023; Lam et al., 2023) have been transferred to ozone forecasting by leveraging very large datasets and models (Bodnar et al., 2024). In weather studies there is work on forecasting at observation stations using these methods, and transferring these methods to forecast ozone at ground-level stations is feasible (Manshausen et al., 2024).

3 Applications of ML methods in atmospheric chemistry modeling

3.1 Background

Global modeling of atmospheric chemistry is a grand computational challenge due to the high dimensionality of coupled chemical species, the nonlinearity and numerical stiffness of solving chemical mechanisms, and interactions with transport on all scales. The inclusion of comprehensive atmospheric chemistry in Earth system models (ESMs), which simulate the interactions between the atmosphere, oceans, land surface, and biosphere, is a priority science frontier (National Research Council (U.S.), 2012). Atmospheric chemical mechanisms are typically implemented in CTMs, which focus on the distribution and chemical evolution of species in the atmosphere. For some applications, chemistry-climate models (CCMs) may also couple chemical processes with climate dynamics, allowing feedback between chemistry and climate. Current atmospheric chemistry models integrate the coupled chemical kinetic equations for mechanism species over model time steps using high-order implicit numerical solvers, but these solvers are computationally expensive (Sandu et al., 1997) and often dominate the cost of an atmospheric simulation (Eastham et al., 2018). Such costs put the inclusion of atmospheric chemistry in tension with other computationally intensive ESM/CCM priorities such as increased spatial resolution and ensemble simulations. The current slowdown in the rate of increase in the speed of computer CPUs – the "end of Moore's law" – underscores the need for computationally efficient approaches (Theis and Wong, 2017).

Chemical solvers in atmospheric models compute the local evolution of species concentrations over a chemical time step that may range from minutes to hours depending on the model (Brasseur and Jacob, 2017). The chemical mechanisms used in regional to global atmospheric models and ESMs typically include $\mathcal{O}(100)$ coupled species with chemical lifetimes ranging from less than a second to much larger than the model time step. High-order implicit solvers can integrate this system of stiff coupled differential equations with high accuracy and fast implementations of these schemes are available, but they are still extremely costly for atmospheric models. Atmospheric models may combat that cost by decreasing the size of the chemical mechanism, breaking down the stiffness of the problem, or using lower-order approximations. However, these methods rarely achieve a speedup of more than a factor of two (Lin et al., 2023; Shen et al., 2020) and sometimes lead to loss of accuracy in the model results. As a consequence, these computational barriers limit the ability for high-resolution simulations, prevent detailed uncertainty analyses, and complicate the coupling of atmospheric chemistry into CCMs/ESMs for long-term climate simulations without significant compute resources.

ML methods could be transformative in this area for both reducing the cost of an atmospheric chemistry simulation and facilitating their incorporation into ESMs. ML methods seem well-suited to replace chemical solvers in atmospheric models because the chemical computation is very repetitive, involving the integration of similar conditions in neighboring grid cells and successive time steps. However, the large number of coupled species brings a "curse of dimensionality" to the problem, and ML methods have no check on error growth, unlike in standard chemical solvers where errors are dampened by the negative response to perturbations (Le Chatelier's principle).

3.2 Progress and State of the Science

Largely, ML methods in atmospheric chemistry modeling currently involve emulating model components to improve model parameterizations, reduce computational bottlenecks, and create simplified, reduced-order models. Here, emulation refers to an ML model reproducing the same calculations as a component of a complex physical or simulated system for a set of inputs. There exists a growing number of studies forecasting ozone on short- (hourly) (Yafouz et al., 2022) and longer-term (Du et al., 2022; Chen et al., 2023) timescales, spanning from city- (Ojha et al., 2021) to regional-level (Ortiz et al., 2021) spatial scales. However, few of these studies have been implemented in operational settings (i.e., within CTMs, CCMs) to offer insight beyond that of traditional model-to-observation comparison methods.

3.2.1 Offline ML and reduced order modeling

Xing et al. (2020) used a hierarchy of ML models containing a CNN and long short-term memory (LSTM) network to predict ozone concentrations from CMAQ model output over 7 d forecast periods. Kuo and Fu (2023) investigated how accurately ML models can learn the ozone-NO_x-VOC chemical relationships in a chemical mechanism and found that their ML model produced distorted NO_x and VOC-limited isopleths when only trained on CMAQ model outputs. Kelp et al. (2020) trained an NN integrator in a photochemical box model, including an encoder/decoder to decrease dimensionality, and a recursive feedback loop over 24 h integration time to control error growth. They found that they could compress the 101-species dimension of their mechanism into 16 features without significant error penalty and avoid error growth within a selected time horizon, though error increases beyond this window. Yang et al. (2024) created an ML surrogate for a low-dimensional (11 species) chemical box model that both compresses the dimensionality of the chemical mechanism and reduces the numerical stiffness of the problem. They achieve numerical stability within a 9 d training window but acknowledge that such an approach may be difficult for more complex and higher-dimensional chemical mechanisms. Liu et al. (2024) employed a Fourier Neural Operator

with time-embedded attention to calculate chemical concentration changes as a learnable time-dependent process. They achieved higher accuracy metrics compared to standard neural operators and U-Nets in simple box model-like simulations.

3.2.2 Online ML within global models

While there is a growing literature on using ML to emulate and improve the representation of atmospheric processes, few have implemented these ML models online within CTMs/ESMs to evaluate their effectiveness. Keller and Evans (2019) created a random forest (RF) integrator for the GEOS-Chem global 3-D CTM driven by re-analyzed meteorological data. They achieved successful short-term simulations but found large error growth after a few weeks. Liu et al. (2022a) developed a gas-phase NN solver for the CMAQ regional CTM over China, combining a standard implicit solver for radicals and oxidants with an ML solver for VOCs. They achieved an order of magnitude speedup over a 1-month simulation but with error growth over remote ocean grid cells. Shen et al. (2022) used an unsupervised ML algorithm (simulated annealing) to create submechanisms of the full chemical mechanism in GEOS-Chem for which they solve the coupled kinetic system only for the fast species in the submechanism. The computational cost of the chemical integration decreased by 50 % and the relative difference in ozone was < 0.5 % in the troposphere and < 0.1 % in the stratosphere over 8-year simulations. Kelp et al. (2022) implemented the low-dimensional "Super-Fast" chemical mechanism in GEOS-Chem using online training of the ML emulator, achieving stable 1-year simulations for ozone prediction with less than 10 % bias compared to the reference and reducing computational cost by a factor of five. However, their ML solver had relatively lower accuracy in pristine marine regimes with lower chemical concentrations. Xia et al. (2024) implemented a self-attention transformer chemical solver online into the WRF-Chem CTM achieving an eight-time speedup over the conventional solver with stable bias metrics for 74 species. Their approach shows promise for accurate predictions of chemical concentrations with low overhead when coupling the ML solver to the CTM, but simulations were only run for 15 d and stability over longer time scales (> 1 year) remains to be seen.

While ML models are typically trained and deployed using Python libraries, integration of these models into CTMs remains limited because CTMs are written in Fortran, which cannot natively call Python. Current solutions include rewriting models in neural Fortran (Keller and Evans, 2019), using the C Foreign Function Interface (CFFI) to create C-style bindings for Python scripts (Kelp et al., 2022; Zhong et al., 2023), or packaging ML models as callable static or dynamic libraries using TorchScript and LibTorch (Xia et al., 2024). Depending on the architecture and complexity of the coupled ML model, all coupling methods result in a speedup over the conventional reference solver (de Burgh-Day and Leeuwenburg, 2023).

3.2.3 ML modeling processes affecting ozone chemistry

A number of ML and data-driven advances have been made for CTM modeling that are separate from creating an ML chemical solver. Wiser et al. (2023) and Wang et al. (2023) created automated chemical mechanism reduction approaches to reduce the high dimensionality of the VOC precursors of ozone and secondary organic aerosol. Sturm and Wexler (2022, 2020) developed methods to enforce mass and stoichiometric conservation rules in outputs from ML emulators. Anderson et al. (2022) used gradient-boosted regression trees to develop a parametrization for the OH radical, a key driver of ozone formation, for CCM models. Similarly, Zhu et al. (2022) trained an ML model on CTM output parameters and satellite observations from OMI to predict urban OH concentrations. Huang and Seinfeld (2022) created an NN-assisted Euler integrator to speed up the iterative computations within an implicit solver routine.

There is a growing literature on ML approaches for bias corrections on existing air quality modeling systems (Neal et al., 2014; Borrego et al., 2011; Silibello et al., 2015). These approaches generally learn the error between the output of a numerical model and some observations and then apply this error correction to the output of the numerical model. Silva et al. (2019) developed an ML parameterization for ozone dry deposition velocities using surface observations that outperformed those within CTMs for certain locations. Similarly, Ivatt and Evans (2020) created an eXtreme Gradient Boosting (XGBoost) model trained on ozone surface observations and data from ozonesonde networks to predict and correct GEOS-Chem model biases. Liu et al. (2022a) developed a NN model to correct surface ozone in the UKESM model, finding that temperature drives biases over Northern Hemisphere continental areas while photolysis rates contribute to global ozone biases. Nowack et al. (2018) used a hierarchy of ML methods to build temperature-based ozone parameterizations for climate model sensitivity simulations. Colombi et al. (2023) used RFs to remove the effect of weather coupled to ozone trends. Gouldsbrough et al. (2024) used a gradient-boosted tree to downscale ozone model output from the EMEP4UK CTM. Ye et al. (2022) used an RF model to identify underlying causes of CTM bias in simulating daily surface ozone variability, finding that CTM underestimates in the dry deposition velocity and cloud optical depth on wet/cloudy days were the primary drivers over China.

Park et al. (2023) created a prototype ML discretization for a one-dimensional horizontal passive scalar advection, an operator component common to all CTMs, and achieved stability and orders of magnitude computational gain relative to the reference when coarse-grained. Sturm et al. (2023) developed a data-driven compression method for chemical trac-

ers within a CTM and advected the compressed representation, achieving a computational gain of 1.5× without loss of accuracy. There have been developments of ML emulators in box models for organic aerosol mechanisms detailing the ML models' accuracy with respect to interactions with ozone (Mouchel-Vallon and Hodzic, 2023; Schreck et al., 2022). The photolysis frequencies used to inform ozone concentrations, calculated from the radiative transfer components of atmospheric models, can themselves be emulated using NNs (Lagerquist et al., 2021) and have the longest relative history of ML emulation for atmospheric modeling (Krasnopolsky et al., 2005, 2008).

The near-term future of integrating ML with 3-D atmospheric chemistry and climate modeling relies on understanding the uncertainties and limitations of ML emulation. Such knowledge is essential for improving or approximating specific chemical parameterizations rather than attempting to replace full-scale, multiscale chemistry simulations. Key priorities include incorporating ML models into CTMs, CCMs, and ESMs, as well as characterizing their behavior over extended time scales (> 1 year). While short-to-seasonal scale emulation may be suitable for forecasting horizons, it offers limited applicability for integrating comprehensive atmospheric chemistry into climate simulations.

4 Applications of AI/ML methods to satellite observations

4.1 Background

Satellite measurements provide detailed information on the spatiotemporal distribution of atmospheric composition and related parameters, such as those associated with surface air quality. Satellite measurements have greater spatial and temporal coverage compared to in-situ observations and they can fill the gaps in those sparse distributions, particularly in remote areas where in-situ observations are not available.

Over the past few decades, multiple satellites have been launched to measure total ozone columns. However, total column measurements cannot be used to provide insight into near-surface ozone because the amount of stratospheric ozone is much larger than the amount of tropospheric ozone. Tropospheric ozone information has been directly retrieved using measurements from nadir-viewing thermal infrared (TIR) sounders, such as the Tropospheric Emission Spectrometer (TES) (Bowman et al., 2002) and the Infrared Atmospheric Sounding Interferometer (IASI) (Boynard et al., 2009), and by combining measurements from both ultraviolet (UV) and visible (VIS) wavelengths by the Tropospheric Emissions: Monitoring of Pollution instrument (Johnson et al., 2018). In addition, the limb-nadir matching method employs stratospheric ozone data from limb-viewing measurements, such as those from the Microwave Limb Sounder (MLS), to derive tropospheric columns from observed total columns (Ziemke et al., 2019). Recently, multispectral satellite approaches, such as IASI and the Global Ozone Monitoring Experiment (GOME) 2 (Cuesta et al., 2018) and TES and the Ozone Monitoring Instrument (OMI) (Colombi et al., 2021), have been implemented to derive tropospheric ozone profiles with increased sensitivity to the lower troposphere.

Nevertheless, satellite observations of ozone are still limited in spatial, temporal, and vertical resolution and are not sufficiently sensitive to ground surface levels. On the other hand, measurements of precursors, such as NO_2 and CH_2O from OMI, GOME-2, the Tropospheric Monitoring Instrument (TROPOMI), and the Ozone Mapping and Profiler Suite, have provided unprecedented information to assess the formation processes and surface concentrations of pollutants such as ozone and aerosols. Despite these advancements, technical challenges remain in accurately assessing near-surface air pollutant concentrations from satellite observations of precursors. ML techniques can be used to fill the gaps in the information available from satellite observations and to improve the estimation of surface air pollutants.

4.2 Progress and State of the Science

ML has been widely used in satellite applications, especially in remote sensing imagery (Maxwell et al., 2018) in the past and is becoming more widely applied to atmospheric composition data. ML has been applied to satellite observations in two main categories: (1) to generate atmospheric concentration retrievals and blend multi-satellite products, and (2) to fill gaps in observational information, including surface concentrations and emissions estimates.

4.2.1 ML models for fast retrievals and multi-satellite blending

Ozone retrieval is the task of estimating ozone profiles from spectrometers on satellites, which measure radiance spectra from the atmosphere. ML-driven retrieval algorithms have emerged as a powerful tool to improve the processing efficiency of atmospheric composition satellite products. Traditional physics-based retrievals, which are based on radiative transfer models (RTMs) and solve their inverse problem, have been widely used to generate satellite profiles of atmospheric composition concentrations – known as level 2 (L2) products – from observed spectral radiances. They consider detailed atmospheric processes to retrieve concentrations, but are computationally expensive. To speed up the retrieval processes, numerical inversion schemes have been replaced by ML algorithms that are trained using RTM inversions. Such an approach has been applied to satellite measurements to retrieve ozone (e.g., Müller et al., 2003), SO_2 (e.g., Li et al., 2022), isoprene (e.g., Wells et al., 2022), and CO_2 (e.g., Xie et al., 2024).

In addition, ML techniques have been used to correct for satellite product bias and blending multiple products. For example, Oak et al. (2024) corrected the Geostationary Environment Spectrometer operational L2 NO_2 vertical column density with a ML model to match more mature TROPOMI observations, while preserving the GEMS data density. Similarly, Balasus et al. (2023) created a blend of TROPOMI and Greenhouse Gases Observing Satellite (GOSAT) methane products obtained by training the ML model to predict differences between TROPOMI and GOSAT co-located observations. Shi et al. (2024) developed an ozone column harmonization method using ConvNeXt (Liu et al., 2022b) to learn a mapping between OMI and TROPOMI, creating a reconstructed ozone column product with the long length of OMI availability and high spatial resolution and accuracy characteristics of TROPOMI. Such bias correction and blending approaches are powerful for providing accurate and consistent datasets for various science applications, for example, emissions inversion.

4.2.2 Fill in gaps in observational information

ML can also be used to fill gaps in observational information, such as supplementing missing data due to clouds to provide a continuous spatiotemporal distribution, and providing surface quantities that cannot be directly measured by satellites. Satellite observations of ozone and its precursors, combined with additional information such as meteorological conditions, land-use, population density, and anthropogenic emission inventories, have been used in NN or RF models to estimate spatiotemporal patterns of surface ozone concentrations at high spatial resolutions in different regions of the world (Di et al., 2017; Wang et al., 2022; Zhu et al., 2022; Kang et al., 2021; Ghahremanloo et al., 2023).

Di et al. (2017) proposed a hybrid NN model using data from OMI, GEOS-Chem CTM outputs, ozone vertical profiles, meteorological variables, land-use terms and other atmospheric compounds to predict daily maximum 8 h average (MDA8) ozone in the continental United States. XGBoost was used by Liu et al. (2020) to predict MDA8 ozone with similar inputs, while Jung et al. (2024) used XGBoost with OMI and MODIS products to estimate MDA8 at 1 km resolution in Taiwan. Ghahremanloo et al. (2023) used a CNN with TROPOMI data as an input to estimate MDA8 in the United States. Among various ML techniques, Zong et al. (2024) concluded that Deep Forests perform better than other tree-based regression models to estimate surface ozone from satellite ozone products. Similar surface concentration estimations based on NN or RF models have been applied to satellite NO_2 products to estimate surface NO_2 concentrations with high spatial resolution (Kim et al., 2021), and to satellite aerosol optical depth measurements to estimate surface $PM_{2.5}$ concentrations (Huang et al., 2021; Xiao et al., 2021) which are useful for exposure estimates.

Emissions estimation using satellite observations of atmospheric composition concentration is another important ML application. ML techniques have been applied to improve the computational efficiency and accuracy of emissions estimation at various scales compared to traditional approaches based on data assimilation and other approaches (Dadheech et al., 2025; Xing et al., 2022; Tu et al., 2023; Li et al., 2024; Bruno et al., 2024).

In addition, ML-based anomaly detection methods pinpoint pollution hotspots, such as urban centers and areas of high industrial activity. For instance, Joyce et al. (2023) developed a deep NN to identify and quantify point source emissions of methane from hyperspectral images from the PRecursore IperSpettrale della Missione Applicativa (PRISMA) satellite with 30 m spatial resolution. ML models can also identify contributions from various emission sources (e.g., traffic, industry, wildfires) (Kang and Im, 2024; Finch et al., 2022; Kurchaba et al., 2023; Rollend et al., 2023).

ML can also be used to characterize key chemical environments and classify each area into different chemical regimes based on satellite observations of pollutants and their precursors. For example, the abundance of OH in urban areas initiates the removal of pollutants, making it a key species to describe the urban chemical environment. Despite its importance, it cannot be measured at the regional scale due to its very short chemical lifetime (Duncan et al., 2024).

These results indicate that combining satellite observations with ML approaches can provide important information for understanding and improving air pollution, including surface ozone and its precursor emissions, which cannot be directly measured from satellite observations. Further progress in this area can be expected through careful evaluation and understanding of the characteristics and quality of satellite products, selection of effective supplementary information, and further development of appropriate ML methods.

5 Challenges and Limitations

In this section, we reflect on some common challenges and limitations of using ML in the context of ozone forecasting, modeling, and observations (Fig. 4). While we describe many challenges which are shared with ML for physical modeling in general, we also highlight challenges specific to ozone modeling with ML in the following sections. In particular, we describe challenges related to the diversity and spatial heterogeneity of ozone monitoring datasets, the difficulty of modeling chemical processes operating at different timescales and with limited data on factors influencing ozone concentrations, as well as detailing challenges more generally applicable to ML for physical modeling. Furthermore, in this section and the next, we propose concrete next steps to make progress on those challenges specific to ozone.

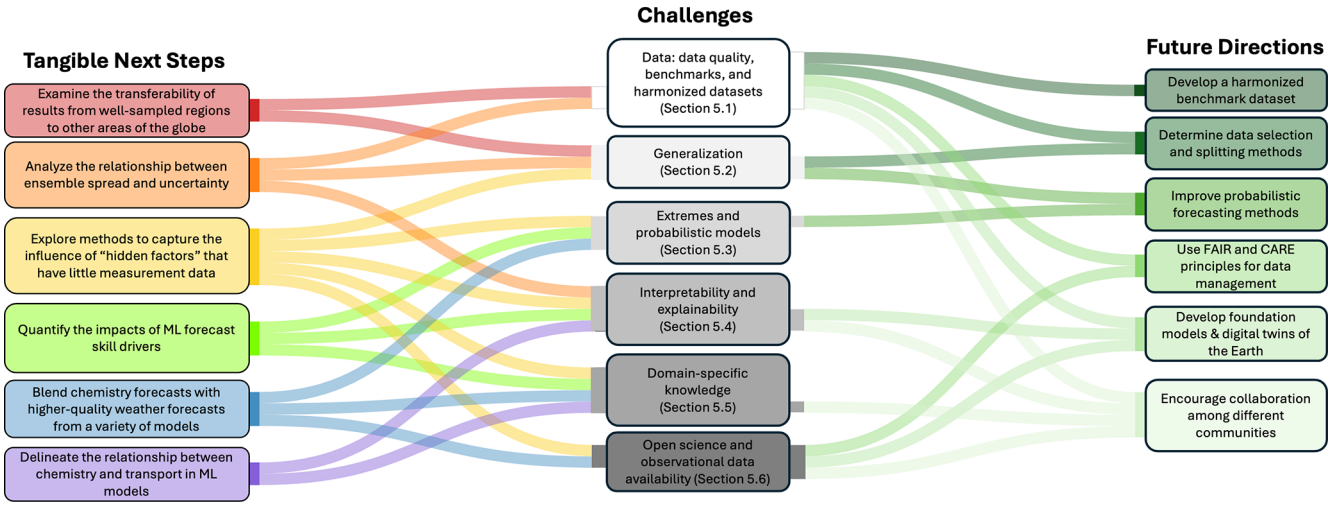

Figure 4. Challenges and future directions described in Sects. 5–6. The middle column represents the categories of challenges described in further detail in the sections listed. The left column lists specific projects that could be undertaken to address the challenges. The right column represents general future directions for the ozone AI/ML modeling community to consider. The lines connect categories of challenges with specific future directions and tasks that could address and resolve those challenges.

5.1 The challenges of data availability and workflow

Central to the success of ML modeling efforts and their utility are the choice of datasets and workflows, i.e., ML model choice and training methods. As noted above, in the field of air pollution and atmospheric composition research, the use of ML is hampered by the absence of benchmark datasets suitable for training different model types with varying sizes and complexity. Such well-defined benchmarks including datasets, training objectives, evaluation scores, and baseline models have been instrumental for the rapid development of ML models in other fields (Dueben et al., 2022). In particular, WeatherBench and WeatherBench2 (Rasp et al., 2020, 2024) have been key factors driving the transformation of ML weather forecasting between 2022 and 2024. A similar dataset to perform the same function for ozone forecasting would allow the robust comparison of different methods, and may guide the field towards more accurate models. Careful curation and data fusion of the TOAR surface ozone database with other relevant datasets might provide a robust and representative benchmark dataset, building on existing work (Betancourt et al., 2021). Ultimately, a lack of sufficient surface observations will impact the study of air quality and downstream impacts on health or vegetation. Therefore improving data coverage over poorly monitored areas remains a priority (Schultz et al., 2017).

The breadth of information available in datasets like TOAR, GHOST and reanalysis products like Copernicus Atmosphere Monitoring Service (CAMS) is vast, but these products provide significant challenges to the development of ML models for ozone due to heterogeneous data formats and lack of succinct documentation that focuses on the use of such data for ML applications. It may be that ML methods can also be used for infilling missing data (e.g. cloud-filtered satellite data, gaps in in-situ observations) for meteorological variables (Li et al., 2023) and ozone (Arroyo et al., 2018; Betancourt et al., 2022). Overall, there is a clear need for a harmonized benchmark dataset(s) for ozone to further enable ML models to be developed. These should (as much as possible) follow the vision outlined by Ebert-Uphoff et al. (2017) and the principles defined by Dueben et al. (2022).

With regards to model choice and development, it is worth noting that, in contrast to CTMs and other methods of simulation, ML models do not a priori require simulation, outputting and aggregating of high-resolution time series data to generate predictions for relevant ozone metrics. Instead, ML models can be trained to directly generate forecasts of these metrics (see Sect. 2). In this regard, it is necessary that ozone ML benchmarks should, where appropriate, include target objectives both for forecasting concentrations and for forecasting (a set of) aggregate ozone metrics. See Fleming et al. (2018) and Lefohn et al. (2018) for a more detailed discussion on relevant ozone metrics.

With regard to ML model training, there is a wide array of data-splitting approaches that can answer subtly related scientific questions. For forecasting it is common practice to divide the data temporally such that the training data completely precedes the testing data, such as using the last few years of a longitudinal dataset for testing. This is commonly recommended in benchmarking studies (Lam et al., 2023; Rasp et al., 2020). We emphasize that while these procedures likely make intuitive sense, they do not match the default setting in ML packages (Schultz et al., 2021), such as scikit-learn (Pedregosa et al., 2011), where the default cross-validation procedure will randomly split over individual data

instances rather than over spatial blocks or temporal blocks. Without using these correct procedures, performance will be overestimated and may not reflect real performance when deployed. In practice, it remains a challenge to carefully define and document the data selection and splitting procedures and adapt them to the scientific problem at hand.

These challenges of data selection and splitting become particularly relevant when looking into climate timescales. Not only can this cause out-of-distribution samples of model input data (for example higher temperatures), but climate change may also affect atmospheric chemical and physical processes so that the mapping between inputs and outputs may drift. This problem is known as "concept drift", where ML predictions become less accurate over time, which can arise from a non-representative training dataset or an ML model that lacks expressiveness, for example by being unable to extrapolate effectively beyond the bounds of the training data. For the latter, tree-based ML models especially are poor with respect to extremes and outliers. Here, model architecture may play a role. Exploring generative AI models, such as Generative Adversarial Networks (GANs) and transformer models, holds promise for the next generation of ML-based atmospheric models. These newer ML architectures can generate more internally consistent dynamics and require less training data than classical CNNs, while also demonstrating improved accuracy and stability over time.

5.2 The challenge of generalization

In addition to appropriate handling of training data, ensuring the trained model is as generally useful as possible, both in and out of sample, remains an enduring challenge. As is common in ML tasks, models trained on data from one geographical region may not necessarily transfer to another region, even when the underlying task and physics remain the same. This limitation often arises from variations in spurious features or unobserved variables specific to each domain, or differences in emissions and climate in different regions. Many approaches in the ML literature seek to improve the performance of ML models across domains, or under domain shifts, which are yet to be used for ozone forecasting (Sagawa et al., 2019), while recent studies suggest that large-scale weather forecasting models may generalize to unseen conditions and perturbations (Hakim and Masanam, 2024). generalization is particularly important for the use of ML models trained in high-data domains and then deployed in low-data domains. In this context, it may be useful to exploit the benefits of probabilistic forecasting (see following section), using models that report uncertainty in unfamiliar domains.

In the context of observational data, ML is increasingly used to derive or enhance geophysical variables from satellite measurements (see Sect. 4). However, a conceptual and practical challenge lies in the appropriate application of these ML-derived products across spatio-temporal scales (Di et al., 2017; Zhu et al., 2022; Tu et al., 2023) or atmospheric regimes (Ghahremanloo et al., 2023). There is a need to evaluate and document the scale-dependence of such products, and to guide their use in downstream modeling and analysis applications accordingly.

5.3 The challenges of extremes and probabilistic models

A relevant application for ozone forecasting and modeling is the study of extremes, including both accurate forecasting and attribution. Extreme ozone concentrations or fluxes can have a large impact on health and vegetation, and are also referred to as low-likelihood high-impact events. By definition, extreme events occur rarely and are hence challenging to accurately represent. There has been some work on approaches to weight extremes more during model training (Steininger et al., 2021). The ability of models to represent extremes is also an important metric that can be used to evaluate the quality of these models. Extremes can thus play a role for uncertainty quantification of the predictive performance of the models (e.g., important for forecast emulators and assessing ML performance on the extremes), where one can distinguish between epistemic (systematic) and aleatoric (statistical) performance. This connects with an increasingly recognized need to evaluate performance in more rigorous and consistent ways: including the development of new benchmark datasets, diagnostics, and metrics (see above). Progress on ML evaluation include causal evaluation (process-oriented approach) and eXplainable AI (xAI), for understanding (in)consistencies of the ML algorithms with physical processes (in other words: whether accurate answers are found for the right reasons). However, such methods are generally only applicable to relatively small-scale ML models. Dynamical tests and counterfactual experiments provide a means to test the credibility of large ML models (Hakim and Masanam, 2024; Baño-Medina et al., 2024).

Forecasting of potential extreme events is particularly challenging because these events are beyond the typical ozone variability, and naturally, extreme events are rarely and infrequently represented in data. For data-driven models this challenge is further exacerbated by (1) the need to forecast not only the presence of threshold exceedances, but also the intensity and duration of extreme events; and (2) ozone extreme events are often related to other anomalous mechanisms, such as heatwaves and wildfires, which are difficult to take into account based on limited extreme information, also due to the fact that ozone responses to these mechanisms are heterogeneous. Although the extreme value theory is widely adopted, its limitations are frequently acknowledged, including the IID (independent and identically distributed) assumption and independence between extreme and non-extreme events. On the other hand, approaches based on probabilistic forecasting may better characterize the uncertainty and likelihood of extreme events. One solution may

be to use metrics and data scenarios to evaluate performance under different types of evaluation scenarios, taking advantage of evaluation metrics in weather forecasting which have been studied extensively. For example, if a key consideration is the ability of a forecasting model to capture extreme events, then metrics that capture relevant performance explicitly on those events should be used. This allows for robust comparison of both the existing and novel models on both traditional metrics and metrics focused on extreme event prediction to more comprehensively evaluate model performance. Evaluation of extreme events is limited in the literature, with recent studies highlighting the lower accuracy of ML models when forecasting spring and summertime ozone concentrations (Leufen et al., 2023; Hickman et al., 2023). In addition to helping with forecasting extreme events, probabilistic forecasting more generally provides a number of advantages compared to the deterministic forecasting methods that are currently more common, as outlined in Bodnar et al. (2024). Furthermore, ML weather forecasting models are increasingly adopting probabilistic and diffusion-based architectures that are able to produce sharp forecasts and uncertainty estimates. This is a promising line of work, however, these ML architectures may be challenging to implement for ozone forecasts due to the uncertainty driven by the meteorological fields themselves.

5.4 The challenge of interpretability and explainability

Interpreting and explaining ML models used to study ozone remains difficult. While these two terms are often used interchangeably, for this article we follow the distinction that interpretability focuses on designing and exploring models that are transparent and have comprehensible internal data transformations, whereas explainability methods focus on post-hoc explanations of how black-box models are working (Rudin, 2019). Models that are directly and trivially interpretable, such as multiple linear regression, are typically not the most performant, and in the high-data regime, the most performant ML models are typically variants on deep NNs that are difficult to interpret or explain. There is some literature that explores whether PINNs provide more interpretability. For example, efforts are underway to enhance the interpretability of ML models in atmospheric sciences by incorporating or diagnosing conservation priorities such as mass and stoichiometry (Sturm and Wexler, 2020, 2022). Additionally, neural operators are being employed to learn the solution operators of ODEs/PDEs from the chemical training data (Liu et al., 2024). However, while incorporating chemistry and physics constraints has been shown to increase interpretability, there is no guarantee that these methods will improve the stability of the ML model over time (Sturm et al., 2023). Often, there is a trade-off between interpretability and ML model accuracy, especially with more complex models (Sengupta et al., 2023). While methods to interpret and explain NNs more generally have been studied widely, mechanistic interpretability of NNs is a challenging task (Nanda et al., 2023), and only a limited range of XAI methods have been tested with ML methods developed for ozone forecasting, often focused on sensitivity approaches which look at the post-hoc explanations where the inputs to models are perturbed to see how predictions change (Ivanovs et al., 2021). Recent studies have investigated the importance of model input parameters through bootstrapping, i.e. random perturbations of individual inputs (Kleinert et al., 2021). Input data perturbation experiments are also possible and informative for very large models as, for example, demonstrated by Hakim and Masanam (2024) for the Pangu-Weather forecast model. Furthermore, ML approaches are increasingly employed to develop end-to-end models that process raw input data (e.g., emissions, meteorological fields) and directly predict outputs such as ozone concentrations. While end-to-end models bypass the challenges of emulating individual components, which are less prone to short-term instabilities and operator splitting issues, they also limit the ability to track uncertainty metrics tied to physical parameters and processes.

5.5 The challenges arising from domain-specific knowledge

Modeling ozone using ML proves challenging due to the multitude of sources driving model error (emissions, chemistry, transport, deposition) and the nonlinear response of ozone to these sources. Parameter tuning an appropriate ozone ML model for a complex, high-dimensional parameter space is possible given large computational resources and adaptive learning on pre-defined metrics. However, such an approach is largely inefficient given that atmospheric chemistry data lies on relatively low-dimensional manifolds with respect to the possible input parameter space. That is, many ozone-related relationships are structured with individual signals often being sparse and low-rank. Here, domain knowledge from atmospheric chemistry can help identify the optimal training dataset and define meaningful loss functions and targeted timescales (Fig. 1) for the ML model problem. In particular, domain knowledge of chemical and physical processes can help explain errors in ML models at short time scales versus long time scales. For example, ML models of atmospheric chemistry tend to predict well fast chemical processes (e.g., seconds to days) but diverge over longer time scales (e.g., months to years) (Kelp et al., 2020). In addition, knowledge of slow chemical processes, such as the role of peroxyacetyl nitrate (PAN) decomposition for ozone formation over polluted and/or remote areas, may help define appropriate training targets for ML models. An emphasis should be placed on emulating chemistry on longer timescales ($>$ 1 year) as issues of long-range stability are more challenging than shorter-term accuracy, and are a necessity for inclusion into CCMs and ESMs.

On the other hand, a heightened focus on domain knowledge may unintentionally limit the potential of ML models. Atmospheric chemists typically leverage well-established relationships of the chemical system, such as NO_x-limited vs. VOC-limited regimes, which are easily uncovered by linear regression or principal component analysis. By invoking such a strong prior assumption, it may impose constraints that hinder an ML model's ability to learn more complex, non-obvious interactions within the data. This bias toward known relationships risks overlooking patterns that could be hidden in the chemical state space that may promote greater accuracy and stability over longer time scales. Striking a balance between leveraging domain expertise and allowing ML models the flexibility to explore complex dynamics is essential for advancing the predictive capability of ozone modeling.

5.6 The challenges of open science and observational data availability

Although an open data infrastructure such as the TOAR-II database gives the impression of low barriers to data access, this might in fact not be true for everyone. Poor internet connectivity from developing countries may limit researchers from retrieving data and subsequently running a computationally demanding model (Blanken et al., 2022; Dwivedi et al., 2022). Furthermore, not all possible data providers agree with sending their data to an open access database, which is one important factor that limits global coverage of surface measurement data. The increasing resolution of satellite products and models is often considered to be an improvement, but the larger data size can complicate the processing and analysis of data for some researchers (Jain et al., 2022). Some data services require a registration and compliance with data use policies, which could conflict with institutional policies of researchers or exacerbate language barriers that non-native English researchers can experience. Finally, whereas advanced APIs can be ideal for technically skilled researchers and allow for reproducible workflows, they might hinder less technical researchers or policy makers that want to explore data sets.

In particular to developing nations, which may not have the economic ability to acquire high-resolution satellite products outside of those freely-available, it is imperative to develop high-quality, globally generalizable solutions to ozone modeling. Data hosting platforms like Google Earth Engine (GEE) enable users to freely access global data relevant for ozone modeling studies, ranging from land-use information from MODIS (Friedl, 2021) to human modification data from VIIRS nighttime lights (Elvidge et al., 2017), Gridded Population of the World (CIESIN, 2018), and more. Recent work by Kazemi Garajeh et al. (2023) investigated the ability to detect spatially resolved ozone pollution trends using time-series Sentinel-5 imagery from GEE, highlighting the quality of spatial distribution and accuracy available an open-source product and platform. This demonstrates the necessity to co-design data services and their hosting platforms to provide efficient and performant access to high-quality, well-documented data.

6 Future Directions

While new developments have been made in DL to utilize the expansive Earth observation data now available (Eyring et al., 2024), challenges remain regarding the quality, interpretability, and complexity of available data. Also, less work has been done to exploit atmospheric composition datasets, where observations are often less dense and more noisy than weather data. Future research and advancements in observational products suitable for ML, including efforts to address uncertainty quantification (e.g., Haynes et al., 2023), will enhance our understanding, facilitate process-based model evaluation (Nowack et al., 2020), and enable actionable science. Accurate forecasts of extrema in short-term surface ozone predictions are essential for protecting human health, while reliable projections of long-term changes in tropospheric ozone abundances are critical for understanding climate change and its impacts. Leveraging causal- and physics-constrained data-driven approaches can enhance trust and interpretability in ML-based modeling efforts (Tesch et al., 2023; Beucler et al., 2024), and combining causal discovery and xAI methods holds potential for advanced process-based evaluation (Iglesias-Suarez et al., 2024). There is a recognized need to evaluate model performance rigorously and consistently, calling for the development of new benchmark datasets, diagnostics, and metrics (Betancourt et al., 2021, 2022), to enable comprehensive evaluation of ML-based ozone modeling techniques. To meet society's needs facing current environmental challenges by providing actionable science and maintaining rapid progress in this field, collaboration among atmospheric composition communities and ML communities is essential.

To thrive, the interdisciplinary ozone modeling and forecasting community requires open knowledge sharing, resources and research cooperation. Research in the domain should adhere to the FAIR principles of Findability, Accessibility, Interoperability and Reusability (Wilkinson et al., 2008) and the CARE principles of Collective, Authority to control, Responsibility and Ethics (Carroll et al., 2021). Availability of data is essential for data-driven approaches and the developed TOAR-II surface ozone database is essential here through its open data policies and its Application Programming Interface (API), which allow for automatic extraction of data.

Future work may focus on foundation models to advance more integrated approaches. These models, trained on extensive datasets in a self-supervised manner Bommasani et al. (2021), have already demonstrated their capability in fields like weather forecasting and climate sci-

ence (Lessig et al., 2023; Nguyen et al., 2023; Bodnar et al., 2024). In the context of tropospheric ozone modeling, foundation models could improve performance by learning from varied datasets, including observational and numerical modeling data (Mukkavilli et al., 2023). These models are capable of handling multiple air pollutants simultaneously and can incorporate meteorological variables, supporting the development of more comprehensive, flexible and potentially robust air quality benchmarks by harmonizing observational data. Their flexible architectures enable training a single model with large-scale resources and then fine-tuning it for multiple tasks, reducing the computational expense of repeated model training (Bommasani et al., 2021).

At present, there are underexplored opportunities to merge the current successes in ML weather and climate model emulation with CTMs and ESMs. Thus far, atmospheric chemistry data have been largely excluded from ML weather and climate applications, as these current supervised learning frameworks are typically non-extensible, requiring retraining of the entire ML model when incorporating new chemical information. In contrast, unsupervised learning model frameworks, such as pre-trained foundation models, can identify patterns in data without explicit labels, offering a new frontier for ingesting and potentially improving ML modeling of atmospheric chemistry. These foundation models can be fine-tuned on CTM data. For example, the Aurora model (Bodnar et al., 2024) is fine-tuned on a subset of six criteria pollutants, including ozone from CAMS (Inness et al., 2019). Fine-tuning ML weather and climate models enables the addition of chemical species to an ML model that is already trained on atmospheric dynamics. This process of fine-tuning, by training specific decoders for new variables, has also recently been carried out for hydrological variables (Lehmann et al., 2025). However, adding chemical species such as VOCs in the absence of emission inputs (which current models do not consider) on the ML weather model's native 6 h forecast time steps likely presents challenges. Greater emphasis is needed on understanding the factors influencing ML model performance with respect to the specific challenges of atmospheric composition research and air quality analysis.

While much progress will likely be made by large coordinated efforts to build comprehensive datasets and foundation models (or fine-tune existing foundation models), progress on important problems specific to ozone may be achieved without requiring large-scale, compute-intensive projects (see Fig. 4). First, investigation of the influence of sparsely observed factors on the skill of ozone forecasting models would be useful. For example, VOCs are not widely measured, and the impact of including VOCs as inputs to an ozone forecasting model could be explored. Second, the capacity of models to predict ozone formation in under-observed regions, after being trained on well-sampled regions, may also provide information about the generalizability of ML models. This work could also inform the importance of largely unobserved variables that influence ozone but differ between regions. More generally, a systematic exploration of the factors that influence machine learning model skill in modeling ozone would be useful for the field. In addition, since much data for model training comes from CTMs, which often have degraded resolution compared to the most accurate weather models, improved model accuracy may be obtained by combining high resolution weather models with chemistry data from lower resolution models. This relates to delineating chemistry and transport in models. Since ML models do not explicitly transport chemical tracers, it may be interesting to explore how ML models perform for tracers with different lifetimes. Furthermore, as probabilistic machine learning methods are established for ozone modeling, analyzing the relationship between ensemble spread and ensemble error could be insightful.

7 Conclusions

While modeling ozone accurately remains a challenging problem across temporal and spatial scales, ML approaches have made progress in a number of areas. As highlighted in this Perspective, ML methods are contributing to research in short-term forecasting, chemistry model emulation, and remote sensing of ozone. Specifically, ML methods are providing increasingly accurate short-term forecasts of ozone at observational stations, and making progress toward providing fast emulators of chemical mechanisms used in chemistry-climate models. In remote sensing, ML methods have shown skill in increasing the efficiency of ozone retrieval, and in making estimates of ozone where there is little satellite coverage.

Similar to many applications of ML to physical modeling, for our field to make progress in modeling real-world ozone faithfully, models should be trained with a synthesis of high-quality observational datasets and appropriate high-quality benchmarks must be compiled to evaluate the skill of different models, and enable the comparison of ML and numerical models. Furthermore, continued work to mitigate the ozone-specific challenges faced by existing ML models is necessary, as highlighted in Sect. 6, which will require close collaboration between domain experts and ML researchers to develop models tailored to the particular challenges. Notably for ozone modeling, recent work illustrates that foundation models, trained on diverse datasets, are capable of skillful atmospheric composition modeling. The paradigm of foundation models represents a significant step forward for composition modeling, enabling an integrated approach across multiple scales and tasks, and building on the success of similar models for weather forecasting. However, it remains an open and important question whether ML models can contribute to improved process-level understanding of drivers of ozone, including quantifying the influence of sparsely observed drivers, and generalize to unseen air pollution and climate scenarios.

As ML continues to transform ozone research and adjacent fields, including weather and climate modeling, the ozone modeling community needs to ensure future research builds on strong foundations. By developing robust benchmarks, building productive cross-disciplinary collaborations, and embracing state-of-the-art techniques, ML-driven ozone research has the potential to not only advance scientific understanding but also deliver actionable benefits for climate resilience and public health.

Appendix A: Glossary

Glossary of terms and abbreviations

Artificial Intelligence (AI)	A software or model that is capable of performing tasks that typically require human intelligence.
Copernicus Atmosphere Monitoring Service (CAMS)	A service by the EU Earth observation programme to provide comprehensive data on atmospheric composition and air quality through satellite and ground-based monitoring.
Chemistry-Climate Model (CCM)	A type of global model focused on the interactions between atmospheric chemistry and climate.
Convolutional Neural Network (CNN)	A type of neural network designed for processing data with grid structure, often used for image processing.
Chemical Transport Model (CTM)	A type of global model designed to simulate the movement and chemical reactions of atmospheric pollutants.
Deep Forest (DF)	A deep learning architecture based on decision trees instead of neural networks.
Deep Learning (DL)	A field of machine learning focused on the development and use of neural networks.
Decision Tree (DT)	A hierarchical supervised learning algorithm, often used to create classification and regression models.
Earth System Model (ESM)	A global model that simulates all aspects of the Earth system, including the interactions between the atmosphere, oceans, land, and biosphere.
Feed-forward Neural Network (FNN)	A basic type of neural network where data move in one direction without feedback loops, often used for data classification and recognition.
Foundation Model (FM)	A machine learning model trained on vast amounts of data, designed to be adapted to a broad range of tasks.
Generative Adversarial Network (GAN)	A type of machine learning technique where two neural networks compete unsupervised to produce the most accurate result.
General Circulation Model (GCM)	A global model that simulates the Earth's atmospheric dynamics and circulation.
Gradient Boosted Decision Tree (GBDT)	An ensemble machine learning technique that uses the results of multiple decision trees to improve accuracy and reduce error of the prediction.
Large Language Model (LLM)	A type of foundation model trained on very large text datasets to understand and generate natural language.
Long Short-Term Memory network (LSTM)	A type of recurrent neural network designed to retain information over longer sequences for longer periods.
Machine Learning (ML)	A field of artificial intelligence dedicated to algorithms and models that can learn and make predictions from the input data without being explicitly programmed to do so.
Neural Network (NN)	A machine learning model designed to process data in a similar way as the human brain.
Physics-Informed Neural Network (PINN)	A type of neural network trained to follow known physical laws.
Random Forest (RF)	An ensemble machine learning algorithm that combines multiple decision trees during the training process to improve prediction accuracy and reduce overfitting.
Recurrent Neural Networks (RNN)	A type of neural network in which data can loop back into the network retaining memory of previous inputs. It is designed for sequential data processing where context is important, such as natural language and time series.
Transformer Model (TM)	A type of deep learning model that converts a given input into a desired output, learning context and meaning. It is used as a foundation model for large language models as an alternative to CNNs and RNNs architectures.
U-Net	A type of convolutional neural network designed for image segmentation and de-noising.
eXplainable AI (xAI)	A type of artificial intelligence that provides the information necessary to understand how a certain output was achieved.

Code and data availability. All code and plotting routines are available at https://github.com/ML4O3/Applications-of-Machine-Learning-and-Artificial-Intelligence-in-Tropospheric-Ozone-Research (ML4O3, 2024; https://doi.org/10.5281/zenodo.17546216, Griffiths, 2025) and TOAR surface ozone data are available at Schultz et al. (2017) (https://doi.org/10.17616/R3FZ0G, re3data.org, 2025).

Author contributions. All authors contributed to the writing and review of the manuscript. SHMH, MK, PTG, FIS, GK, KD, KZ, MGS and EAP led the writing and preparation of the MS. SHMH led Sect. 2, MK Sect. 3, KD, EAP and KM led Sect. 4, GK Sect. 5 and FIS Sect. 6.

Competing interests. At least one of the (co-)authors is a guest member of the editorial board of *Atmospheric Chemistry and Physics* for the special issue "Tropospheric Ozone Assessment Report Phase II (TOAR-II) Community Special Issue (ACP/AMT/BG/ESSD/GMD inter-journal SI)". The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. The authors acknowledge the TOAR community for their support of this work.

Financial support. SHMH acknowledges funding from EPSRC via the AI4ER CDT at the University of Cambridge (EP/S022961/1). PTG acknowledges the Environmental Geochemical Cycle Research Group, Earth Surface System Research Center, Japan Agency For Marine Science and Technology for support as an External Researcher, and the Global Atmospheric Chemistry Section, Earth System Division, National Institute for Environmental Studies, Tsukuba, Japan for support as a visiting scientist in 2024. PTG and ATA were financially supported by NERC through NCAS (R8/H12/83/003). Part of this work was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the NASA. We acknowledge the support of the National Aeronautics and Space Administration (NASA) Atmospheric Composition: Aura Science Team Program (19-AURAST19-0044), Atmospheric Composition Modeling and Analysis Program (22-ACMAP22-0013), NASA Earth Science U.S. Participating Investigator program (22-EUSPI22-0005). ZL was acknowledges the National Natural Science Foundation of China, grant number 42307140. MGS acknowledges funding by the European Commission under grant ERC-AdvG-787576. KLC was supported by NOAA cooperative agreement (no. NA22OAR4320151). JJW acknowledges National Aeronautics and Space Administration (NASA) grants no. NNX16AQ30G and no. 80NSSC23K0930. DEC acknowledges financial support from Underwriters Laboratories, Inc.

Review statement. This paper was edited by Juan Antonio Añel and reviewed by Brian Henn and one anonymous referee.

References

Agapiou, A.: Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine© applications, International Journal of Digital Earth, 10, 85–102, https://doi.org/10.1080/17538947.2016.1250829, 2017.

Alari, A., Schwarz, L., Zabrocki, L., Le Nir, G., Chaix, B., and Benmarhnia, T.: The effects of an air quality alert program on premature mortality: A difference-in-differences evaluation in the region of Paris, Environment International, 156, 106583, https://doi.org/10.1016/j.envint.2021.106583, 2021.

Anderson, D. C., Follette-Cook, M. B., Strode, S. A., Nicely, J. M., Liu, J., Ivatt, P. D., and Duncan, B. N.: A machine learning methodology for the generation of a parameterization of the hydroxyl radical, Geosci. Model Dev., 15, 6341–6358, https://doi.org/10.5194/gmd-15-6341-2022, 2022.

Archibald, A. T., Neu, J. L., Elshorbany, Y. F., Cooper, O. R., Young, P. J., Akiyoshi, H., Cox, R. A., Coyle, M., Derwent, R. G., Deushi, M., Finco, A., Frost, G. J., Galbally, I. E., Gerosa, G., Granier, C., Griffiths, P. T., Hossaini, R., Hu, L., Jöckel, P., Josse, B., Lin, M. Y., Mertens, M., Morgenstern, O., Naja, M., Naik, V., Oltmans, S., Plummer, D. A., Revell, L. E., Saiz-Lopez, A., Saxena, P., Shin, Y. M., Shahid, I., Shallcross, D., Tilmes, S., Trickl, T., Wallington, T. J., Wang, T., Worden, H. M., and Zeng, G.: Tropospheric Ozone Assessment Report: Critical review of changes in the tropospheric ozone burden and budget from 1960–2100, Elementa: Science of the Anthropocene, 8, 034, https://doi.org/10.1525/elementa.2020.034, 2020.

Arroyo, A., Herrero, A., Tricio, V., Corchado, E., and Woźniak, M.: Neural models for imputation of missing ozone data in air-quality datasets, Complexity, 1, 7238015, https://doi.org/10.1155/2018/7238015, 2018.

Balasus, N., Jacob, D. J., Lorente, A., Maasakkers, J. D., Parker, R. J., Boesch, H., Chen, Z., Kelp, M. M., Nesser, H., and Varon, D. J.: A blended TROPOMI+GOSAT satellite data product for atmospheric methane using machine learning to correct retrieval biases, Atmos. Meas. Tech., 16, 3787–3807, https://doi.org/10.5194/amt-16-3787-2023, 2023.

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47–55, https://doi.org/10.1038/nature14956, 2015.

Baño-Medina, J., Sengupta, A., Doyle, J. D., Reynolds, C. A., Watson-Parris, D., and Monache, L. D.: Are AI weather models learning atmospheric physics? A sensitivity analysis of cyclone Xynthia, Research Square [preprint], https://doi.org/10.21203/rs.3.rs-5356949/v1, 2024.

Bell, M. L., Zanobetti, A., and Dominici, F.: Who is More Affected by Ozone Pollution? A Systematic Review and

Meta-Analysis, American Journal of Epidemiology, 180, 15, https://doi.org/10.1093/aje/kwu115, 2014.

Betancourt, C., Stomberg, T., Roscher, R., Schultz, M. G., and Stadtler, S.: AQ-Bench: a benchmark dataset for machine learning on global air quality metrics, Earth Syst. Sci. Data, 13, 3013–3033, https://doi.org/10.5194/essd-13-3013-2021, 2021.

Betancourt, C., Stomberg, T. T., Edrich, A.-K., Patnala, A., Schultz, M. G., Roscher, R., Kowalski, J., and Stadtler, S.: Global, high-resolution mapping of tropospheric ozone – explainable machine learning and impact of uncertainties, Geosci. Model Dev., 15, 4331–4354, https://doi.org/10.5194/gmd-15-4331-2022, 2022.

Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., Yu, S., Rasp, S., Ahmed, F., O'Gorman, P., and Neelin, J.: Climate-invariant machine learning, Science Advances, 10, eadj7250, https://doi.org/10.1126/sciadv.adj7250, 2024.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, https://doi.org/10.1038/s41586-023-06185-3, 2023.

Blanken, P. D., Brunet, D., Dominguez, C., Goursaud Oger, S., Hussain, S., Jain, M., Koren, G., Mu, Y., Ray, P., Saxena, P., Sonwani, S., and Sur, D.: Atmospheric sciences perspectives on integrated, coordinated, open, networked (ICON) science, Earth and Space Science, https://doi.org/10.1029/2021EA002204, 2022.

Bodnar, C., Bruinsma, W., Lucic, A., Stanley, M., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J., Dong, H., Vaughan, A., and Gupta, J.: Aurora: a foundation model of the atmosphere, arXiv [preprint], https://doi.org/10.48550/arXiv.2405.13063, 2024.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., and others: On the opportunities and risks of foundation models, arXiv [preprint], https://doi.org/10.48550/arXiv.2108.07258, 2021.

Borrego, C., Monteiro, A., Pay, M., Ribeiro, I., Miranda, A., Basart, S., and Baldasano, J.: How bias-correction can improve air quality forecasts over Portugal, Atmospheric Environment, 45, 6629–6641, 2011.

Bowman, K. W., Steck, T., Worden, H. M., Worden, J., Clough, S., and Rodgers, C.: Capturing time and vertical variability of tropospheric ozone: A study using TES nadir retrievals, Journal of Geophysical Research: Atmospheres, 107, ACH 21–1–ACH 21–11, https://doi.org/10.1029/2002JD002150, 2002.

Boynard, A., Clerbaux, C., Coheur, P.-F., Hurtmans, D., Turquety, S., George, M., Hadji-Lazaro, J., Keim, C., and Meyer-Arnek, J.: Measurements of total and tropospheric ozone from IASI: comparison with correlative satellite, ground-based and ozonesonde observations, Atmos. Chem. Phys., 9, 6255–6271, https://doi.org/10.5194/acp-9-6255-2009, 2009.

Brasseur, G. P. and Jacob, D. J.: Modeling of Atmospheric Chemistry, Cambridge University Press, Cambridge, ISBN 978-1-107-14696-9, https://doi.org/10.1017/9781316544754, 2017.

Bruno, J. H., Jervis, D., Varon, D. J., and Jacob, D. J.: U-Plume: automated algorithm for plume detection and source quantification by satellite point-source imagers, Atmos. Meas. Tech., 17, 2625–2636, https://doi.org/10.5194/amt-17-2625-2024, 2024.

Buonocore, J., Robinson, L., Hammitt, J., and O'Keeffe, L.: Estimating the potential health benefits of air quality warnings, Risk Analysis, 41, 645–660, 2021.

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., and Stall, S.: Operationalizing the CARE and FAIR principles for indigenous data futures, Scientific Data, 8, 108, https://doi.org/10.1038/s41597-021-00892-0, 2021.

Casciaro, G., Cavaiola, M., and Mazzino, A.: Calibrating the CAMS European multi-model air quality forecasts for regional air pollution monitoring, Atmospheric Environment, 287, 119259, https://doi.org/10.1016/j.atmosenv.2022.119259, 2022.

Chaloulakou, A., Saisana, M., and Spyrellis, N.: Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens, Science of the Total Environment, 313, 1–13, 2003.

Chen, B., Wang, Y., Huang, J., Zhao, L., Chen, R., Song, Z., and Hu, J.: Estimation of near-surface ozone concentration and analysis of main weather situation in China based on machine learning model and Himawari-8 TOAR data, Science of The Total Environment, 864, 160928, https://doi.org/10.1016/j.scitotenv.2022.160928, 2023.

Chen, J., Shen, H., Li, X., Li, T., and Wei, Y.: Ground-level ozone estimation based on geo-intelligent machine learning by fusing in-situ observations, remote sensing data, and model simulation data, International Journal of Applied Earth Observation and Geoinformation, 112, 102955, https://doi.org/10.1016/j.jag.2022.102955, 2022.

Cheng, M., Fang, F., Navon, I. M., Zheng, J., Tang, X., Zhu, J., and Pain, C.: Spatio-Temporal Hourly and Daily Ozone Forecasting in China Using a Hybrid Machine Learning Model: Autoencoder and Generative Adversarial Networks, Journal of Advances in Modeling Earth Systems, 14, e2021MS002806, https://doi.org/10.1029/2021MS002806, 2022.

Christiansen, A., Mickley, L. J., Liu, J., Oman, L. D., and Hu, L.: Multidecadal increases in global tropospheric ozone derived from ozonesonde and surface site observations: can models reproduce ozone trends?, Atmos. Chem. Phys., 22, 14751–14782, https://doi.org/10.5194/acp-22-14751-2022, 2022.

Center For International Earth Science Information Network – CIESIN – Columbia University: Gridded population of the world, version 4 (GPWv4): Population count, revision 11, https://doi.org/10.7927/H4PN93PB, 2018.

Clifton, O. E., Fiore, A. M., Massman, W. J., Baublitz, C. B., Coyle, M., Emberson, L., Fares, S., Farmer, D. K., Gentine, P., Gerosa, G., Guenther, A. B., Helmig, D., Lombardozzi, D. L., Munger, J. W., Patton, E. G., Pusede, S. E., Schwede, D. B., Silva, S. J., Sörgel, M., Steiner, A. L., and Tai, A. P. K.: Dry Deposition of Ozone Over Land: Processes, Measurement, and Modeling, Reviews of Geophysics, 58, e2019RG000670, https://doi.org/10.1029/2019RG000670, 2020.

Cobourn, W. G., Dolcine, L., French, M., and Hubbard, M. C.: A comparison of nonlinear regression and neural network models for ground-level ozone forecasting, Journal of the Air & Waste Management Association, 50, 1999–2009, 2000.

Colombi, N., Miyazaki, K., Bowman, K. W., Neu, J. L., and Jacob, D. J.: A new methodology for inferring surface ozone from multi-spectral satellite measurements, Environmental Research Letters, 16, 105005, https://doi.org/10.1088/1748-9326/ac243d, 2021.

Colombi, N. K., Jacob, D. J., Yang, L. H., Zhai, S., Shah, V., Grange, S. K., Yantosca, R. M., Kim, S., and Liao, H.: Why is ozone in South Korea and the Seoul metropolitan area so

high and increasing?, Atmos. Chem. Phys., 23, 4031–4044, https://doi.org/10.5194/acp-23-4031-2023, 2023.

Comrie, A.: Comparing neural networks and regression models for ozone forecasting, Journal of the Air & Waste Management Association, 47, 653–663, 1997.

Cooper, O. R., Schultz, M. G., Schröder, S., Chang, K.-L., Gaudel, A., Benítez, G. C., Cuevas, E., Fröhlich, M., Galbally, I. E., Molloy, S., Kubistin, D., Lu, X., McClure-Begley, A., Nédélec, P., O'Brien, J., Oltmans, S. J., Petropavlovskikh, I., Ries, L., Senik, I., Sjöberg, K., Solberg, S., Spain, G. T., Spangl, W., Steinbacher, M., Tarasick, D., Thouret, V., and Xu, X.: Multi-decadal surface ozone trends at globally distributed remote locations, Elementa: Science of the Anthropocene, 8, 23, https://doi.org/10.1525/elementa.420, 2020.

Cuesta, J., Kanaya, Y., Takigawa, M., Dufour, G., Eremenko, M., Foret, G., Miyazaki, K., and Beekmann, M.: Transboundary ozone pollution across East Asia: daily evolution and photochemical production analysed by IASI + GOME2 multispectral satellite observations and models, Atmos. Chem. Phys., 18, 9499–9525, https://doi.org/10.5194/acp-18-9499-2018, 2018.

Dadheech, N., He, T.-L., and Turner, A. J.: High-resolution greenhouse gas flux inversions using a machine learning surrogate model for atmospheric transport, Atmos. Chem. Phys., 25, 5159–5174, https://doi.org/10.5194/acp-25-5159-2025, 2025.

de Burgh-Day, C. O. and Leeuwenburg, T.: Machine learning for numerical weather and climate modelling: a review, Geosci. Model Dev., 16, 6433–6477, https://doi.org/10.5194/gmd-16-6433-2023, 2023.

Di, Q., Rowland, S., Koutrakis, P., and Schwartz, J.: A hybrid model for spatially and temporally resolved ozone exposures in the continental United States, Journal of the Air & Waste Management Association, 67, 39–52, https://doi.org/10.1080/10962247.2016.1200159, 2017.

Du, J., Qiao, F., Lu, P., and Yu, L.: Forecasting ground-level ozone concentration levels using machine learning, Resources, Conservation and Recycling, 184, 106380, https://doi.org/10.1016/j.resconrec.2022.106380, 2022.

Dueben, P., Schultz, M., Chantry, M., Gagne, D., Hall, D., and McGovern, A.: Challenges and benchmark datasets for machine learning in the atmospheric sciences: Definition, status, and outlook, Artificial Intelligence for the Earth Systems, 1, e210002, https://doi.org/10.1175/AIES-D-21-0002.1, 2022.

Duncan, B. N., Anderson, D. C., Fiore, A. M., Joiner, J., Krotkov, N. A., Li, C., Millet, D. B., Nicely, J. M., Oman, L. D., St. Clair, J. M., Shutter, J. D., Souri, A. H., Strode, S. A., Weir, B., Wolfe, G. M., Worden, H. M., and Zhu, Q.: Opinion: Beyond global means – novel space-based approaches to indirectly constrain the concentrations of and trends and variations in the tropospheric hydroxyl radical (OH), Atmos. Chem. Phys., 24, 13001–13023, https://doi.org/10.5194/acp-24-13001-2024, 2024.

Dwivedi, D., Santos, A. L. D., Barnard, M. A., Crimmins, T. M., Malhotra, A., Rod, K. A., Aho, K. S., Bell, S. M., Bomfim, B., Brearley, F. Q., Cadillo-Quiroz, H., Chen, J., Gough, C. M., Graham, E. B., Hakkenberg, C. R., Haygood, L., Koren, G., Lilleskov, E. A., Meredith, L. K., Naeher, S., Nickerson, Z. L., Pourret, O., Song, H.-S., Stahl, M., Tas, N., Vargas, R., and Weintraub-Leff, S.: Biogeosciences perspectives on integrated, coordinated, open, networked (ICON) science, Earth and Space Science, https://doi.org/10.1029/2021EA002119, 2022.

Eastham, S. D., Long, M. S., Keller, C. A., Lundgren, E., Yantosca, R. M., Zhuang, J., Li, C., Lee, C. J., Yannetti, M., Auer, B. M., Clune, T. L., Kouatchou, J., Putman, W. M., Thompson, M. A., Trayanov, A. L., Molod, A. M., Martin, R. V., and Jacob, D. J.: GEOS-Chem High Performance (GCHP v11-02c): a next-generation implementation of the GEOS-Chem chemical transport model for massively parallel applications, Geosci. Model Dev., 11, 2941–2953, https://doi.org/10.5194/gmd-11-2941-2018, 2018.

Ebert-Uphoff, I., Thompson, D. R., Demir, I., Gel, Y. R., Karpatne, A., Guereque, M., Kumar, V., Cabral-Cano, E., and Smyth, P.: A vision for the development of benchmarks to bridge geoscience and data science, 17th International Workshop on Climate informatics, https://par.nsf.gov/biblio/10143795-vision-development-benchmarks-bridge-geoscience-data-science (last access: 6 November 2025), 2017.

Elvidge, K. B., D., C., and Ghosh, T.: VIIRS night-time lights, International Journal of Remote Sensing, 38, 5860–5879, https://doi.org/10.3390/rs13050922, 2017.

EPA, U.: Integrated science assessment (ISA) for ozone and related photochemical oxidants (final report, Apr 2020), 2020.

Eslami, E., Choi, Y., Lops, Y., and Sayeed, A.: A real-time hourly ozone prediction system using deep convolutional neural network, Neural Computing and Applications, 32, 8783–8797, 2020.

Eyring, V., Collins, W., Gentine, P., Barnes, E., Barreiro, M., Beucler, T., Bocquet, M., Bretherton, C., Christensen, H., Dagon, K., and Gagne, D.: Pushing the frontiers in climate modelling and analysis with machine learning, Nature Climate Change, 14 1–13, 2024.

Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., and Wang, J.: Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation, Atmospheric Environment, 107, 118–128, 2015.

Finch, D. P., Palmer, P. I., and Zhang, T.: Automated detection of atmospheric NO_2 plumes from satellite data: a tool to help infer anthropogenic combustion emissions, Atmos. Meas. Tech., 15, 721–733, https://doi.org/10.5194/amt-15-721-2022, 2022.

Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., Doherty, R. M., Horowitz, L. W., MacKenzie, I. A., Sanderson, M. G., Shindell, D. T., Stevenson, D. S., Szopa, S., Van Dingenen, R., Zeng, G., Atherton, C., Bergmann, D., Bey, I., Carmichael, G., Collins, W. J., Duncan, B. N., Faluvegi, G., Folberth, G., Gauss, M., Gong, S., Hauglustaine, D., Holloway, T., Isaksen, I. S. A., Jacob, D. J., Jonson, J. E., Kaminski, J. W., Keating, T. J., Lupu, A., Marmer, E., Montanaro, V., Park, R. J., Pitari, G., Pringle, K. J., Pyle, J. A., Schroeder, S., Vivanco, M. G., Wind, P., Wojcik, G., Wu, S., and Zuber, A.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, Journal of Geophysical Research, 114, D04301, https://doi.org/10.1029/2008JD010816, 2009.

Fiore, A. M., Hancock, S. E., Lamarque, J.-F., Correa, G. P., Chang, K.-L., Ru, M., Cooper, O., Gaudel, A., Polvani, L. M., Sauvage, B., and Ziemke, J. R.: Understanding recent tropospheric ozone trends in the context of large internal variability: a new perspective from chemistry-climate model ensembles, Environmental Research: Climate, 1, 025008, https://doi.org/10.1088/2752-5295/ac9cc2, 2022.

Fleming, Z., Doherty, R., Von Schneidemesser, E., Malley, C., Cooper, O., Pinto, J., Colette, A., Xu, X., Simpson, D., Schultz, M., and Lefohn, A.: Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health, Elem. Sci. Anth., 6, 12, https://doi.org/10.1525/elementa.273, 2018.

Friedl, S.-M.: MCD12Q1 MODIS/terra+aqua land cover type yearly L3 global 500m SIN grid V006, NASA Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/MODIS/MCD12Q1.006, 2021.

Gardner, M. and Dorling, S.: Statistical surface ozone models: an improved methodology to account for non-linear behaviour, Atmospheric Environment, 34, 21–34, 2000.

Ghahremanloo, M., Choi, Y., and Lops, Y.: Deep learning mapping of surface MDA8 ozone: The impact of predictor variables on ozone levels over the contiguous United States, Environmental Pollution, 326, 121508, https://doi.org/10.1016/j.envpol.2023.121508, 2023.

Gouldsbrough, L., Hossaini, R., Eastoe, E., Young, P. J., and Vieno, M.: A machine learning approach to downscale EMEP4UK: analysis of UK ozone variability and trends, Atmos. Chem. Phys., 24, 3163–3196, https://doi.org/10.5194/acp-24-3163-2024, 2024.

Griffiths, P.: ML4O3/Applications-of-Machine-Learning-and-Artificial-Intelligence-in-Tropospheric-Ozone-Research: GMD, Zenodo [code], https://doi.org/10.5281/zenodo.17546216, 2025.

Gulev, S., Thorne, P., Ahn, J., Dentener, F., Domingues, C., Gerland, S., Gong, D., Kaufman, D., Nnamchi, H., Quaas, J., Rivera, J., Sathyendranath, S., Smith, S., Trewin, B., von Schuckmann, K., and Vose, R.: Changing state of the climate system, in: Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, UK and New York, NY, USA, 287–422, https://doi.org/10.1017/9781009157896.004, , 2021.

Hahm, Y. and Yoon, H.: The impact of air pollution alert services on respiratory diseases: generalized additive modeling study in South Korea, Environmental Research Letters, 16, 064048, https://doi.org/10.1088/1748-9326/ac002f, 2021.

Hakim, G. and Masanam, S.: Dynamical tests of a deep-learning weather prediction model, Artificial Intelligence for the Earth Systems, 3, https://doi.org/10.1175/AIES-D-23-0090.1, 2024.

Han, W., He, T.-L., Jiang, Z., Zhu, R., Jones, D., Miyazaki, K., and Shen, Y.: The Capability of Deep Learning Model to Predict Ozone Across Continents in China, the United States and Europe, Geophysical Research Letters, 50, e2023GL104928, https://doi.org/10.1029/2023GL104928, 2023.

Haynes, K., Lagerquist, R., McGraw, M., Musgrave, K., and Ebert-Uphoff, I.: Creating and evaluating uncertainty estimates with neural networks for environmental-science applications, Artificial Intelligence for the Earth Systems, 2, 220061, https://doi.org/10.1175/AIES-D-22-0061.1, 2023.

Health Effects Institute: State of Global Air 2024: A Special Report on Global Exposure to Air Pollution and Its Health Impacts, With a Focus on Children's Health, Special Report, Health Effects Institute, Boston, MA, 2024.

Heo, J.-S. and Kim, D.-S.: A new method of ozone forecasting using fuzzy expert and neural network systems, Science of The Total Environment, 325, 221–237, https://doi.org/10.1016/j.scitotenv.2003.11.009, 2004.

Hickman, S., Griffiths, P., Nowack, P., and Archibald, A.: Short-term forecasting of ozone air pollution across Europe with transformers, Environmental Data Science, 2, e43, https://doi.org/10.1017/eds.2023.37, 2023.

Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, Neural Networks, 2, 359–366, https://doi.org/10.1016/0893-6080(89)90020-8, 1989.

Huang, C., Hu, J., Xue, T., Xu, H., and Wang, M.: High-Resolution Spatiotemporal Modeling for Ambient PM2.5 Exposure Assessment in China from 2013 to 2019, Environmental Science & Technology, 55, 2152–2162, https://doi.org/10.1021/acs.est.0c05815, 2021.

Huang, Y. and Seinfeld, J. H.: A Neural Network-Assisted Euler Integrator for Stiff Kinetics in Atmospheric Chemistry, Environmental Science & Technology, 56, 4676–4685, https://doi.org/10.1021/acs.est.1c07648, 2022.

Iglesias-Suarez, F., Gentine, P., Solino-Fernandez, B., Beucler, T., Pritchard, M., Runge, J., and Eyring, V.: Causally-informed deep learning to improve climate models and projections, Journal of Geophysical Research: Atmospheres, 129, e2023JD039202, https://doi.org/10.1029/2023JD039202, 2024.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys., 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

Ivanovs, M., Kadikis, R., and Ozols, K.: Perturbation-based methods for explaining deep neural networks: A survey, Pattern Recognition Letters, 150, 228–234, 2021.

Ivatt, P. D. and Evans, M. J.: Improving the prediction of an atmospheric chemistry transport model using gradient-boosted regression trees, Atmos. Chem. Phys., 20, 8063–8082, https://doi.org/10.5194/acp-20-8063-2020, 2020.

Jain, S., Kaur, N., Verma, S., Kavita, Hosen, A. S. M. S., and Sehgal, S. S.: Use of Machine Learning in Air Pollution Research: A Bibliographic Perspective, Electronics, 11, 3621, https://doi.org/10.3390/electronics11213621, 2022.

Johnson, M. S., Liu, X., Zoogman, P., Sullivan, J., Newchurch, M. J., Kuang, S., Leblanc, T., and McGee, T.: Evaluation of potential sources of a priori ozone profiles for TEMPO tropospheric ozone retrievals, Atmos. Meas. Tech., 11, 3457–3477, https://doi.org/10.5194/amt-11-3457-2018, 2018.

Joyce, P., Ruiz Villena, C., Huang, Y., Webb, A., Gloor, M., Wagner, F. H., Chipperfield, M. P., Barrio Guilló, R., Wilson, C., and Boesch, H.: Using a deep neural network to detect methane point sources and quantify emissions from PRISMA hyperspectral satellite images, Atmos. Meas. Tech., 16, 2627–2640, https://doi.org/10.5194/amt-16-2627-2023, 2023.

Jung, C.-R., Chen, W., Chen, W.-T., Su, S.-H., Chen, B.-T., Chang, L., and Hwang, B.-F.: A machine learning model for estimating daily maximum 8-hour average ozone concentrations using OMI

and MODIS products, Atmospheric Environment, 331, 120587, https://doi.org/10.1016/j.atmosenv.2024.120587, 2024.

Kang, Y. and Im, J.: Mitigating underestimation of fire emissions from the Advanced Himawari Imager: A machine learning and multi-satellite ensemble approach, International Journal of Applied Earth Observation and Geoinformation, 128, 103784, https://doi.org/10.1016/j.jag.2024.103784, 2024.

Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C.-K., and Kim, S.: Estimation of surface-level NO_2 and O_3 concentrations using TROPOMI data and machine learning over East Asia, Environmental Pollution, 288, 117711, https://doi.org/10.1016/j.envpol.2021.117711, 2021.

Kazemi Garajeh, M., Laneve, G., Rezaei, H., Sadeghnejad, M., Mohamadzadeh, N., and Salmani, B.: Monitoring trends of CO, NO_2, SO_2, and O_3 pollutants using time-series sentinel-5 images based on google earth engine, Pollutants, 3, 255–279, https://doi.org/10.3390/pollutants3020019, 2023.

Keller, C. A. and Evans, M. J.: Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10, Geosci. Model Dev., 12, 1209–1225, https://doi.org/10.5194/gmd-12-1209-2019, 2019.

Kelp, M. M., Jacob, D. J., Kutz, J. N., Marshall, J. D., and Tessum, C. W.: Toward Stable, General Machine-Learned Models of the Atmospheric Chemical System, Journal of Geophysical Research: Atmospheres, 125, e2020JD032759, https://doi.org/10.1029/2020JD032759, 2020.

Kelp, M. M., Jacob, D. J., Lin, H., and Sulprizio, M. P.: An Online-Learned Neural Network Chemical Solver for Stable Long-Term Global Simulations of Atmospheric Chemistry, Journal of Advances in Modeling Earth Systems, 14, e2021MS002926, https://doi.org/10.1029/2021MS002926, 2022.

Kim, M., Brunner, D., and Kuhlmann, G.: Importance of satellite observations for high-resolution mapping of near-surface NO_2 by machine learning, Remote Sensing of Environment, 264, 112573, https://doi.org/10.1016/j.rse.2021.112573, 2021.

Kleinert, F., Leufen, L. H., and Schultz, M. G.: IntelliO3-ts v1.0: a neural network approach to predict near-surface ozone concentrations in Germany, Geosci. Model Dev., 14, 1–25, https://doi.org/10.5194/gmd-14-1-2021, 2021.

Kolehmainen, M., Martikainen, H., and Ruuskanen, J.: Neural networks and periodic components used in air quality forecasting, Atmospheric Environment, 35, 815–825, 2001.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Chalikov, D. V.: New Approach to Calculation of Atmospheric Model Physics: Accurate and Fast Neural Network Emulation of Longwave Radiation in a Climate Model, Monthly Weather Review, 133, 1370–1383, https://doi.org/10.1175/MWR2923.1, 2005.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Belochitski, A. A.: Decadal Climate Simulations Using Accurate and Fast Neural Network Emulation of Full, Longwave and Shortwave, Radiation, Monthly Weather Review, 136, 3683–3695, https://doi.org/10.1175/2008MWR2385.1, 2008.

Kuo, C.-P. and Fu, J. S.: Ozone response modeling to NOx and VOC emissions: Examining machine learning models, Environment International, 176, 107 969, https://doi.org/10.1016/j.envint.2023.107969, 2023.

Kurchaba, S., van Vliet, J., Verbeek, F. J., and Veenman, C. J.: Anomalous Anomalous NO emitting ship detection with TROPOMI satellite data and machine learning, Remote Sensing of Environment, 297, 113 761, https://doi.org/10.1016/j.rse.2023.113761, 2023.

Lagerquist, R., Turner, D., Ebert-Uphoff, I., Stewart, J., and Hagerty, V.: Using Deep Learning to Emulate and Accelerate a Radiative Transfer Model, Journal of Atmospheric and Oceanic Technology, 38, 1673–1696, https://doi.org/10.1175/JTECH-D-21-0007.1, 2021.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, https://doi.org/10.1126/science.adi2336, 2023.

Lefohn, A., Malley, C., Smith, L., Wells, B., Hazucha, M., Simon, H., Naik, V., Mills, G., Schultz, M., Paoletti, E., and De Marco, A.: Tropospheric ozone assessment report: Global ozone metrics for climate change, human health, and crop/ecosystem research, Elem. Sci. Anth., 6, 27, https://doi.org/10.1525/elementa.279, 2018.

Lehmann, F., Ozdemir, F., Soja, B., Hoefler, T., Mishra, S., and Schemm, S.: Finetuning a Weather Foundation Model with Lightweight Decoders for Unseen Physical Processes, arXiv [preprint], https://doi.org/10.48550/arXiv.2506.19088, 2025.

Lelieveld, J. and Dentener, F. J.: What controls tropospheric ozone?, Journal of Geophysical Research: Atmospheres, 105, 3531–3551, https://doi.org/10.1029/1999JD901011, 2000.

Lessig, C., Luise, I., Gong, B., Langguth, M., Stadtler, S., and Schultz, M.: AtmoRep: A stochastic model of atmosphere dynamics using large scale representation learning, arXiv [preprint], https://doi.org/10.48550/arXiv.2308.13280, 2023.

Leufen, L. H., Kleinert, F., and Schultz, M. G.: O3ResNet: A Deep Learning – Based Forecast System to Predict Local Ground-Level Daily Maximum 8-Hour Average Ozone in Rural and Suburban Environments, Artificial Intelligence for the Earth Systems, 2, https://doi.org/10.1175/AIES-D-22-0085.1, 2023.

Li, C., Joiner, J., Liu, F., Krotkov, N. A., Fioletov, V., and McLinden, C.: A new machine-learning-based analysis for improving satellite-retrieved atmospheric composition data: OMI SO2 as an example, Atmos. Meas. Tech., 15, 5497–5514, https://doi.org/10.5194/amt-15-5497-2022, 2022.

Li, C., Ren, X., and Zhao, G.: Machine-learning-based imputation method for filling missing values in ground meteorological observation data, Algorithms, 16, 422, https://doi.org/10.3390/a16090422, 2023.

Li, Y., Liu, S., Bashiri Khuzestani, R., Huang, K., and Bao, F.: Emission-Based Machine Learning Approach for Large-Scale Estimates of Black Carbon in China, Remote Sensing, 16, 837, https://doi.org/10.3390/rs16050837, 2024.

Lin, H., Long, M. S., Sander, R., Sandu, A., Yantosca, R. M., Estrada, L. A., Shen, L., and Jacob, D. J.: An Adaptive Auto-Reduction Solver for Speeding Up Integration of Chemical Kinetics in Atmospheric Chemistry Models: Implementation and Evaluation in the Kinetic Pre-Processor (KPP) Version 3.0.0, Journal of Advances in Modeling Earth Systems, 15, e2022MS003293, https://doi.org/10.1029/2022MS003293, 2023.

Liu, R., Ma, Z., Liu, Y., Shao, Y., Zhao, W., and Bi, J.: Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach, Environment International,

142, 105823, https://doi.org/10.1016/j.envint.2020.105823, 2020.

Liu, Z., Doherty, R. M., Wild, O., O'Connor, F. M., and Turnock, S. T.: Correcting ozone biases in a global chemistry–climate model: implications for future ozone, Atmos. Chem. Phys., 22, 12543–12557, https://doi.org/10.5194/acp-22-12543-2022, 2022a.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S.: A ConvNet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11976–11986, https://openaccess.thecvf.com/content/CVPR2022/html/Liu_A_ConvNet_for_the_2020s_CVPR_2022_paper.html (last access: 6 November 2025), 2022b.

Liu, Z.-S., Clusius, P., and Boy, M.: Neural Network Emulator for Atmospheric Chemical ODE, arXiv [preprint], https://doi.org/10.48550/arXiv.2408.01829, 2024.

Lu, X., Zhang, L., Chen, Y., Zhou, M., Zheng, B., Li, K., Liu, Y., Lin, J., Fu, T.-M., and Zhang, Q.: Exploring 2016–2017 surface ozone pollution over China: source contributions and meteorological influences, Atmos. Chem. Phys., 19, 8339–8361, https://doi.org/10.5194/acp-19-8339-2019, 2019.

Malashock, D. A., Delang, M. N., Becker, J. S., Serre, M. L., West, J. J., Chang, K.-L., Cooper, O. R., and Anenberg, S. C.: Global trends in ozone concentration and attributable mortality for urban, peri-urban, and rural areas between 2000 and 2019: a modelling study, The Lancet Planetary Health, 6, e958–e967, https://doi.org/10.1016/S2542-5196(22)00260-1, 2022.

Malley, C. S., Henze, D. K., Kuylenstierna, J. C., Vallack, H. W., Davila, Y., Anenberg, S. C., Turner, M. C., and Ashmore, M. R.: Updated Global Estimates of Respiratory Mortality in Adults ≥30Years of Age Attributable to Long-Term Ozone Exposure, Environmental Health Perspectives, 125, 087021, https://doi.org/10.1289/EHP1390, 2017.

Manshausen, P., Cohen, Y., Pathak, J., Pritchard, M., Garg, P., Mardani, M., Kashinath, K., Byrne, S., and Brenowitz, N.: Generative Data Assimilation of Sparse Weather Station Observations at Kilometer Scales, arXiv [preprint], https://doi.org/10.48550/arXiv.2406.16947, 2024.

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., Denier van der Gon, H. A. C., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, Á., van Velthoven, P., van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geosci. Model Dev., 8, 2777–2813, https://doi.org/10.5194/gmd-8-2777-2015, 2015.

Maxwell, A. E., Warner, T. A., and Fang, F.: Implementation of machine-learning classification in remote sensing: an applied review, International Journal of Remote Sensing, 39, 2784–2817, https://doi.org/10.1080/01431161.2018.1433343, 2018.

Mills, G., Sharps, K., Simpson, D., Pleijel, H., Broberg, M., Uddling, J., Jaramillo, F., Davies, W. J., Dentener, F., Van den Berg, M., Agrawal, M., Agrawal, S. B., Ainsworth, E. A., Büker, P., Emberson, L., Feng, Z., Harmens, H., Hayes, F., Kobayashi, K., Paoletti, E., and Van Dingenen, R.: Ozone pollution will compromise efforts to increase global wheat production, Global Change Biology, 24, 3560–3574, https://doi.org/10.1111/gcb.14157, 2018.

Miyazaki, K., Bowman, K. W., Yumimoto, K., Walker, T., and Sudo, K.: Evaluation of a multi-model, multi-constituent assimilation framework for tropospheric chemical reanalysis, Atmos. Chem. Phys., 20, 931–967, https://doi.org/10.5194/acp-20-931-2020, 2020.

ML4O3: ML4O3/Applications-of-Machine-Learning-and-Artificial-Intelligence-in-Tropospheric-Ozone-Research, Github [code], https://github.com/ML4O3/Applications-of-Machine-Learning-and-Artificial-Intelligence-in-Tropospheric-Ozone-Research (last access: 6 November 2025), 2024.

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K. S., Mills, G. E., Stevenson, D. S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., and Williams, M. L.: Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer, Atmos. Chem. Phys., 15, 8889–8973, https://doi.org/10.5194/acp-15-8889-2015, 2015.

Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bekki, S., Butchart, N., Chipperfield, M. P., Deushi, M., Dhomse, S. S., Garcia, R. R., Hardiman, S. C., Horowitz, L. W., Jöckel, P., Josse, B., Kinnison, D., Lin, M., Mancini, E., Manyin, M. E., Marchand, M., Marécal, V., Michou, M., Oman, L. D., Pitari, G., Plummer, D. A., Revell, L. E., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T. Y., Tilmes, S., Yamashita, Y., Yoshida, K., and Zeng, G.: Review of the global models used within phase 1 of the Chemistry–Climate Model Initiative (CCMI), Geosci. Model Dev., 10, 639–671, https://doi.org/10.5194/gmd-10-639-2017, 2017.

Mouchel-Vallon, C. and Hodzic, A.: Toward Emulating an Explicit Organic Chemistry Mechanism With Random Forest Models, Journal of Geophysical Research: Atmospheres, 128, e2022JD038227, https://doi.org/10.1029/2022JD038227, 2023.

Mukkavilli, S. K., Civitarese, D. S., Schmude, J., Jakubik, J., Jones, A., Nguyen, N., Phillips, C., Roy, S., Singh, S., Watson, C., Ganti, R., Hamann, H., Nair, U., Ramachandran, R., and Weldemariam, K.: AI Foundation Models for Weather and Climate: Applications, Design, and Implementation, arXiv [preprint], https://doi.org/10.48550/arXiv.2309.10808, 2023.

Müller, M. D., Kaifel, A. K., Weber, M., Tellmann, S., Burrows, J. P., and Loyola, D.: Ozone profile retrieval from Global Ozone Monitoring Experiment (GOME) data using a neural network approach (Neural Network Ozone Retrieval System (NNORSY)), Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD002784, 2003.

Nanda, N., Chan, L., Lieberum, T., Smith, J., and Steinhardt, J.: Progress measures for grokking via mechanistic interpretability, arXiv [preprint], https://doi.org/10.48550/arXiv.2301.05217, 2023.

National Research Council (U.S.): A National Strategy for Advancing Climate Modeling, National Academies Press, ISBN 978-0-309-25977-4, https://doi.org/10.17226/13430, 2012.

Neal, L., Agnew, P., Moseley, S., Ordóñez, C., Savage, N., and Tilbee, M.: Application of a statistical post-processing technique to a gridded, operational, air quality forecast, Atmospheric Environment, 98, 385–393, 2014.

Neu, J. L., Flury, T., Manney, G. L., Santee, M. L., Livesey, N. J., and Worden, J.: Tropospheric ozone variations governed by changes in stratospheric circulation, Nature Geoscience, 7, 340–344, https://doi.org/10.1038/ngeo2138, 2014.

Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., and Grover, A.: ClimaX: A foundation model for weather and climate, arXiv [preprint], https://doi.org/10.48550/arXiv.2301.10343, 2023.

Nowack, P., Braesicke, P., Haigh, J., Abraham, N. L., Pyle, J., and Voulgarakis, A.: Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations, Environmental Research Letters, 13, 104016, https://doi.org/10.1088/1748-9326/aae2be, 2018.

Nowack, P., Runge, J., Eyring, V., and Haigh, J.: Causal networks for climate model evaluation and constrained projections, Nature Communications, 11, 1415, https://doi.org/10.1038/s41467-020-15195-y, 2020.

Nunnari, G., Nucifora, A., and Randieri, C.: The application of neural techniques to the modelling of time-series of atmospheric pollution data, Ecological Modelling, 111, 187–205, 1998.

Oak, Y. J., Jacob, D. J., Balasus, N., Yang, L. H., Chong, H., Park, J., Lee, H., Lee, G. T., Ha, E. S., Park, R. J., Kwon, H.-A., and Kim, J.: A bias-corrected GEMS geostationary satellite product for nitrogen dioxide using machine learning to enforce consistency with the TROPOMI satellite instrument, Atmos. Meas. Tech., 17, 5147–5159, https://doi.org/10.5194/amt-17-5147-2024, 2024.

Ojha, N., Girach, I., Sharma, K., Sharma, A., Singh, N., and Gunthe, S. S.: Exploring the potential of machine learning for simulations of urban ozone variability, Scientific Reports, 11, 22513, https://doi.org/10.1038/s41598-021-01824-z, 2021.

Ortiz, E. Y., Keller, C. A., Cardenas, B., Sáenz, H. E., Wakabayashi, K., Seddon, J., Castillero, E., Retama, A., Hernandez, L. A., and Grajales, F.: Combination of GEOS-CF model with Machine Learning as a tool for forecasting regional pollution in Bogotá, in: 2021 Congreso Colombiano y Conferencia Internacional de Calidad de Aire y Salud Pública (CASAP), 1–4, https://doi.org/10.1109/CASAP54985.2021.9703381, 2021.

Park, M., Zheng, Z., Riemer, N., and Tessum, C. W.: Learned 1-D passive scalar advection to accelerate chemical transport modeling: a case study with GEOS-FP horizontal wind fields, arXiv [preprint], https://doi.org/10.48550/arXiv.2309.11035, 2023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., Lam, R., and Willson, M.: GenCast: Diffusion-based ensemble forecasting for medium-range weather, arXiv [preprint], https://doi.org/10.48550/arXiv.2312.15796, 2024.

Rasp, S., Pritchard, M. S., and Gentine, P.: Deep learning to represent subgrid processes in climate models, Proceedings of the National Academy of Sciences, 115, 9684–9689, https://doi.org/10.1073/pnas.1810286115, 2018.

Rasp, S., Dueben, P., Scher, S., Weyn, J., Mouatadid, S., and Thuerey, N.: WeatherBench: a benchmark data set for data-driven weather forecasting, Journal of Advances in Modeling Earth Systems, 12, e2020MS002203, https://doi.org/10.1073/pnas.1810286115, 2020.

Rasp, S., Hoyer, S., Merose, A., Langmore, I., Battaglia, P., Russell, T., Sanchez-Gonzalez, A., Yang, V., Carver, R., Agrawal, S., and Chantry, M.: WeatherBench 2: A benchmark for the next generation of data-driven global weather models, Journal of Advances in Modeling Earth Systems, 16, e2023MS004019, https://doi.org/10.1029/2023MS004019, 2024.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

re3data.org: TOAR Surface Observation Database; editing status 2025-03-12; re3data.org – Registry of Research Data Repositories [data set], https://doi.org/10.17616/R3FZ0G, 2025.

Rollend, D., Foster, K., Kott, T. M., Mocharla, R., Muñoz, R., Fendley, N., Ashcraft, C., Willard, F., Reilly, E. P., and Hughes, M.: Machine learning for activity-based road transportation emissions estimation, Environmental Data Science, 2, e38, https://doi.org/10.1017/eds.2023.32, 2023.

Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence, 1, 206–215, 2019.

Saberian, S., Heyes, A., and Rivers, N.: Alerts work! Air quality warnings and cycling, Resource and Energy Economics, 49, 165–185, 2017.

Sagawa, S., Koh, P., Hashimoto, T., and Liang, P.: Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization, arXiv [preprint], https://doi.org/10.48550/arXiv.1911.08731, 2019.

Sandu, A., Verwer, J., Blom, J., Spee, E., Carmichael, G., and Potra, F.: Benchmarking stiff ode solvers for atmospheric chemistry problems II: Rosenbrock solvers, Atmospheric Environment, 31, 3459–3472, https://doi.org/10.1016/S1352-2310(97)83212-8, 1997.

Savage, N. H., Agnew, P., Davis, L. S., Ordóñez, C., Thorpe, R., Johnson, C. E., O'Connor, F. M., and Dalvi, M.: Air quality modelling using the Met Office Unified Model (AQUM OS24-26): model description and initial evaluation, Geosci. Model Dev., 6, 353–372, https://doi.org/10.5194/gmd-6-353-2013, 2013.

Sayeed, A., Choi, Y., Eslami, E., Jung, J., Lops, Y., Salman, A. K., Lee, J.-B., Park, H.-J., and Choi, M.-H.: A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance, Scientific Reports, 11, 10891, https://doi.org/10.1038/s41598-021-90446-6, 2021.

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., and Vondracek, J.: A rigorous inter-comparison of ground-level ozone predictions, Atmospheric Environment, 37, 3237–3253, 2003.

Schreck, J. S., Becker, C., Gagne, D. J., Lawrence, K., Wang, S., Mouchel-Vallon, C., Choi, J., and Hodzic, A.: Neural Network Emulation of the Formation of Organic Aerosols Based on the Explicit GECKO-A Chemistry Model, Journal

of Advances in Modeling Earth Systems, 14, e2021MS002974, https://doi.org/10.1029/2021MS002974, 2022.

Schultz, M. G., Schröder, S., Lyapina, O., and Cooper, O. R. e. a.: Tropospheric Ozone Assessment Report: Database and metrics data of global surface ozone observations, Elementa: Science of the Anthropocene, 5, 58, https://doi.org/10.1525/elementa.244, 2017.

Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., and Stadtler, S.: Can deep learning beat numerical weather prediction?, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379, 20200097, https://doi.org/10.1098/rsta.2020.0097, 2021.

Sengupta, P., Zhang, Y., Maharjan, S., and Eliassen, F.: Balancing Explainability-Accuracy of Complex Models, arXiv [preprint], https://doi.org/10.48550/arXiv.2305.14098, 2023.

Shah, V., Jacob, D. J., Dang, R., Lamsal, L. N., Strode, S. A., Steenrod, S. D., Boersma, K. F., Eastham, S. D., Fritz, T. M., Thompson, C., Peischl, J., Bourgeois, I., Pollack, I. B., Nault, B. A., Cohen, R. C., Campuzano-Jost, P., Jimenez, J. L., Andersen, S. T., Carpenter, L. J., Sherwen, T., and Evans, M. J.: Nitrogen oxides in the free troposphere: implications for tropospheric oxidants and the interpretation of satellite NO_2 measurements, Atmos. Chem. Phys., 23, 1227–1257, https://doi.org/10.5194/acp-23-1227-2023, 2023.

Shen, L., Jacob, D. J., Santillana, M., Wang, X., and Chen, W.: An adaptive method for speeding up the numerical integration of chemical mechanisms in atmospheric chemistry models: application to GEOS-Chem version 12.0.0, Geosci. Model Dev., 13, 2475–2486, https://doi.org/10.5194/gmd-13-2475-2020, 2020.

Shen, L., Jacob, D. J., Santillana, M., Bates, K., Zhuang, J., and Chen, W.: A machine-learning-guided adaptive algorithm to reduce the computational cost of integrating kinetics in global atmospheric chemistry models: application to GEOS-Chem versions 12.0.0 and 12.9.1, Geosci. Model Dev., 15, 1677–1687, https://doi.org/10.5194/gmd-15-1677-2022, 2022.

Shi, C., Zhang, Z., Xiong, S., Chen, W., Zhang, W., Zhang, Q., and Wang, X.: Harmonizing atmospheric ozone column concentrations over the Tibetan Plateau from 2005 to 2022 using OMI and Sentinel-5P TROPOMI: A deep learning approach, International Journal of Applied Earth Observation and Geoinformation, 129, 103808, https://doi.org/10.1016/j.jag.2024.103808, 2024.

Silibello, C., D'Allura, A., Finardi, S., Bolignano, A., and Sozzi, R.: Application of bias adjustment techniques to improve air quality forecasts, Atmospheric Pollution Research, 6, 928–938, 2015.

Silva, S. J., Heald, C. L., Ravela, S., Mammarella, I., and Munger, J. W.: A Deep Learning Parameterization for Ozone Dry Deposition Velocities, Geophysical Research Letters, 46, 983–989, https://doi.org/10.1029/2018GL081049, 2019.

Skeie, R. B., Myhre, G., Hodnebrog, O., Cameron-Smith, P. J., Deushi, M., Hegglin, M. I., Horowitz, L. W., Kramer, R. J., Michou, M., Mills, M. J., Olivié, D. J. L., Connor, F. M. O., Paynter, D., Samset, B. H., Sellar, A., Shindell, D., Takemura, T., Tilmes, S., and Wu, T.: Historical total ozone radiative forcing derived from CMIP6 simulations, npj Climate and Atmospheric Science, 3, 1–10, https://doi.org/10.1038/s41612-020-00131-0, 2020.

Steininger, M., Kobs, K., Davidson, P., Krause, A., and Hotho, A.: Density-based weighting for imbalanced regression, Machine Learning, 110, 2187–2211, 2021.

Sturm, P. O. and Wexler, A. S.: A mass- and energy-conserving framework for using machine learning to speed computations: a photochemistry example, Geosci. Model Dev., 13, 4435–4442, https://doi.org/10.5194/gmd-13-4435-2020, 2020.

Sturm, P. O. and Wexler, A. S.: Conservation laws in a neural network architecture: enforcing the atom balance of a Julia-based photochemical model (v0.2.0), Geosci. Model Dev., 15, 3417–3431, https://doi.org/10.5194/gmd-15-3417-2022, 2022.

Sturm, P. O., Manders, A., Janssen, R., Segers, A., Wexler, A. S., and Lin, H. X.: Advecting Superspecies: Efficiently Modeling Transport of Organic Aerosol With a Mass-Conserving Dimensionality Reduction Method, Journal of Advances in Modeling Earth Systems, 15, e2022MS003235, https://doi.org/10.1029/2022MS003235, 2023.

Tesch, T., Kollet, S., and Garcke, J.: Causal deep learning models for studying the Earth system, Geosci. Model Dev., 16, 2149–2166, https://doi.org/10.5194/gmd-16-2149-2023, 2023.

Theis, T. N. and Wong, H.-S. P.: The End of Moore's Law: A New Beginning for Information Technology, Computing in Science & Engineering, 19, 41–50, https://doi.org/10.1109/MCSE.2017.29, 2017.

Tu, Q., Hase, F., Chen, Z., Schneider, M., García, O., Khosrawi, F., Chen, S., Blumenstock, T., Liu, F., Qin, K., Cohen, J., He, Q., Lin, S., Jiang, H., and Fang, D.: Estimation of NO2 emission strengths over Riyadh and Madrid from space from a combination of wind-assigned anomalies and a machine learning technique, Atmos. Meas. Tech., 16, 2237–2262, https://doi.org/10.5194/amt-16-2237-2023, 2023.

Wang, S., Schmidt, J. A., Baidar, S., Coburn, S., Dix, B., Koenig, T. K., Apel, E., Bowdalo, D., Campos, T. L., Eloranta, E., Evans, M. J., DiGangi, J. P., Zondlo, M. A., Gao, R.-S., Haggerty, J. A., Hall, S. R., Hornbrook, R. S., Jacob, D., Morley, B., Pierce, B., Reeves, M., Romashkin, P., ter Schure, A., and Volkamer, R.: Active and widespread halogen chemistry in the tropical and subtropical free troposphere, Proceedings of the National Academy of Sciences, 112, 9281–9286, https://doi.org/10.1073/pnas.1505142112, 2015.

Wang, W., Liu, X., Bi, J., and Liu, Y.: A machine learning model to estimate ground-level ozone concentrations in California using TROPOMI data and high-resolution meteorology, Environment International, 158, 106917, https://doi.org/10.1016/j.envint.2021.106917, 2022.

Wang, Z., Couvidat, F., and Sartelet, K.: Implementation of a parallel reduction algorithm in the GENerator of reduced Organic Aerosol mechanisms (GENOA v2.0): Application to multiple monoterpene aerosol precursors, Journal of Aerosol Science, 174, 106248, https://doi.org/10.1016/j.jaerosci.2023.106248, 2023.

Wells, K. C., Millet, D. B., Payne, V. H., Vigouroux, C., Aquino, C. a. B., De Mazière, M., de Gouw, J. A., Graus, M., Kurosu, T., Warneke, C., and Wisthaler, A.: Next-Generation Isoprene Measurements From Space: Detecting Daily Variability at High Resolution, Journal of Geophysical Research: Atmospheres, 127, e2021JD036181, https://doi.org/10.1029/2021JD036181, 2022.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P.,

Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Nature Scientific Data, 3, 1–9, 2008.

Wiser, F., Place, B. K., Sen, S., Pye, H. O. T., Yang, B., Westervelt, D. M., Henze, D. K., Fiore, A. M., and McNeill, V. F.: AMORE-Isoprene v1.0: a new reduced mechanism for gas-phase isoprene oxidation, Geosci. Model Dev., 16, 1801–1821, https://doi.org/10.5194/gmd-16-1801-2023, 2023.

Xia, Z., Zhao, C., Du, Q., Yang, Z., Zhang, M., and Qiao, L.: Advancing Photochemistry Simulation in WRF-Chem V4.0: Artificial Intelligence PhotoChemistry (AIPC) Scheme with Multi-Head Self-Attention Algorithm, ESS Open Archive [preprint], https://www.authorea.com/users/816476/articles/1217166-advancing-photochemistry-simulation-in-wrf-chem-v4-0-artificial-intelligence-photochemistry-aipc-scheme-with-multi-head-self-attention-algorithm (last access: 6 November 2025), 2024.

Xiao, Q., Geng, G., Cheng, J., Liang, F., Li, R., Meng, X., Xue, T., Huang, X., Kan, H., Zhang, Q., and He, K.: Evaluation of gap-filling approaches in satellite-based daily $PM_{2.5}$ prediction models, Atmospheric Environment, 244, 117921, https://doi.org/10.1016/j.atmosenv.2020.117921, 2021.

Xie, F., Ren, T., Zhao, C., Wen, Y., Gu, Y., Zhou, M., Wang, P., Shiomi, K., and Morino, I.: Fast retrieval of XCO2 over east Asia based on Orbiting Carbon Observatory-2 (OCO-2) spectral measurements, Atmos. Meas. Tech., 17, 3949–3967, https://doi.org/10.5194/amt-17-3949-2024, 2024.

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

Xing, J., Li, S., Zheng, S., Liu, C., Wang, X., Huang, L., Song, G., He, Y., Wang, S., Sahu, S. K., Zhang, J., Bian, J., Zhu, Y., Liu, T.-Y., and Hao, J.: Rapid Inference of Nitrogen Oxide Emissions Based on a Top-Down Method with a Physically Informed Variational Autoencoder, Environmental Science & Technology, 56, 9903–9914, https://doi.org/10.1021/acs.est.1c08337, 2022.

Yafouz, A., AlDahoul, N., Birima, A. H., Ahmed, A. N., Sherif, M., Sefelnasr, A., Allawi, M. F., and Elshafie, A.: Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction, Alexandria Engineering Journal, 61, 4607–4622, https://doi.org/10.1016/j.aej.2021.10.021, 2022.

Yang, X., Guo, L., Zheng, Z., Riemer, N., and Tessum, C. W.: Atmospheric chemistry surrogate modeling with sparse identification of nonlinear dynamics, arXiv [preprint], https://doi.org/10.48550/arXiv.2401.06108, 2024.

Ye, X., Wang, X., and Zhang, L.: Diagnosing the Model Bias in Simulating Daily Surface Ozone Variability Using a Machine Learning Method: The Effects of Dry Deposition and Cloud Optical Depth, Environmental Science & Technology, 56, 16665–16675, https://doi.org/10.1021/acs.est.2c05712, 2022.

Yi, J. and Prybutok, V.: A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area, Environmental Pollution, 92, 349–357, 1996.

Young, P. J., Naik, V., Fiore, A. M., Gaudel, A., Guo, J., Lin, M. Y., Neu, J. L., Parrish, D. D., Rieder, H. E., Schnell, J. L., Tilmes, S., Wild, O., Zhang, L., Ziemke, J., Brandt, J., Delcloo, A., Doherty, R. M., Geels, C., Hegglin, M. I., Hu, L., Im, U., Kumar, R., Luhar, A., Murray, L., Plummer, D., Rodriguez, J., Saiz-Lopez, A., Schultz, M. G., Woodhouse, M. T., and Zeng, G.: Tropospheric Ozone Assessment Report: Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends, Elementa: Science of the Anthropocene, 6, 10, https://doi.org/10.1525/elementa.265, 2018.

Zhang, Y., West, J. J., Emmons, L. K., Flemming, J., Jonson, J. E., Lund, M. T., Sekiya, T., Sudo, K., Gaudel, A., Chang, K.-L., Nédélec, P., and Thouret, V.: Contributions of World Regions to the Global Tropospheric Ozone Burden Change From 1980 to 2010, Geophysical Research Letters, 48, e2020GL089184, https://doi.org/10.1029/2020GL089184, 2021.

Zhong, X., Ma, Z., Yao, Y., Xu, L., Wu, Y., and Wang, Z.: WRF–ML v1.0: a bridge between WRF v4.3 and machine learning parameterizations and its application to atmospheric radiative transfer, Geosci. Model Dev., 16, 199–209, https://doi.org/10.5194/gmd-16-199-2023, 2023.

Zhu, Q., Laughner, J. L., and Cohen, R. C.: Combining Machine Learning and Satellite Observations to Predict Spatial and Temporal Variation of near Surface OH in North American Cities, Environmental Science & Technology, 56, 7362–7371, https://doi.org/10.1021/acs.est.1c05636, 2022.

Ziemke, J. R., Oman, L. D., Strode, S. A., Douglass, A. R., Olsen, M. A., McPeters, R. D., Bhartia, P. K., Froidevaux, L., Labow, G. J., Witte, J. C., Thompson, A. M., Haffner, D. P., Kramarova, N. A., Frith, S. M., Huang, L.-K., Jaross, G. R., Seftor, C. J., Deland, M. T., and Taylor, S. L.: Trends in global tropospheric ozone inferred from a composite record of TOMS/OMI/MLS/OMPS satellite measurements and the MERRA-2 GMI simulation , Atmos. Chem. Phys., 19, 3257–3269, https://doi.org/10.5194/acp-19-3257-2019, 2019.

Zong, M., Song, T., Zhang, Y., Feng, Y., and Fan, S.: A Deep Forest Algorithm Based on TropOMI Satellite Data to Estimate Near-Ground Ozone Concentration, Atmosphere, 15, 1020, https://doi.org/10.3390/atmos15091020, 2024.