



# Applications of Machine Learning and Artificial Intelligence in Tropospheric Ozone Research

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

<sup>1</sup>Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK

<sup>2</sup>Doerr School of Sustainability, Stanford University, CA, USA

<sup>3</sup>National Centre for Atmospheric Science, Cambridge University

<sup>4</sup>Department of Computer Science, University of Oxford

<sup>5</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

<sup>6</sup>Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, the Netherlands

<sup>7</sup>Predictia Intelligent Data Solutions S.L., Santander, Spain

<sup>8</sup>Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany

<sup>9</sup>NOAA Chemical Sciences Laboratory, Boulder, CO, USA

<sup>10</sup>Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder,

<sup>11</sup>School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK

<sup>12</sup>School of Chemistry, University of Leicester, Leicester, UK

<sup>13</sup>Civil and Environmental Engineering, Duke University, NC, USA

<sup>14</sup>Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill, NC, USA

<sup>15</sup>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

\*These authors contributed equally to this work.

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

## 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.

5 We highlight the rapid progress made in the field since Phase 1 of the Tropospheric Ozone Assessment Report, 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. Despite these advances, many challenges in the field remain, including the quality of data, benchmarks, and limited model generalisation and explainability. This review provides a comprehensive synthesis of recent advancements, highlights critical challenges, and proposes actionable pathways to further advance ML  
10 applications in ozone research. Achieving this potential will require close collaborations across atmospheric chemistry, ML and computational science, aimed at addressing key challenges such as the development of global benchmark datasets and robust, explainable models.



## 1 Introduction

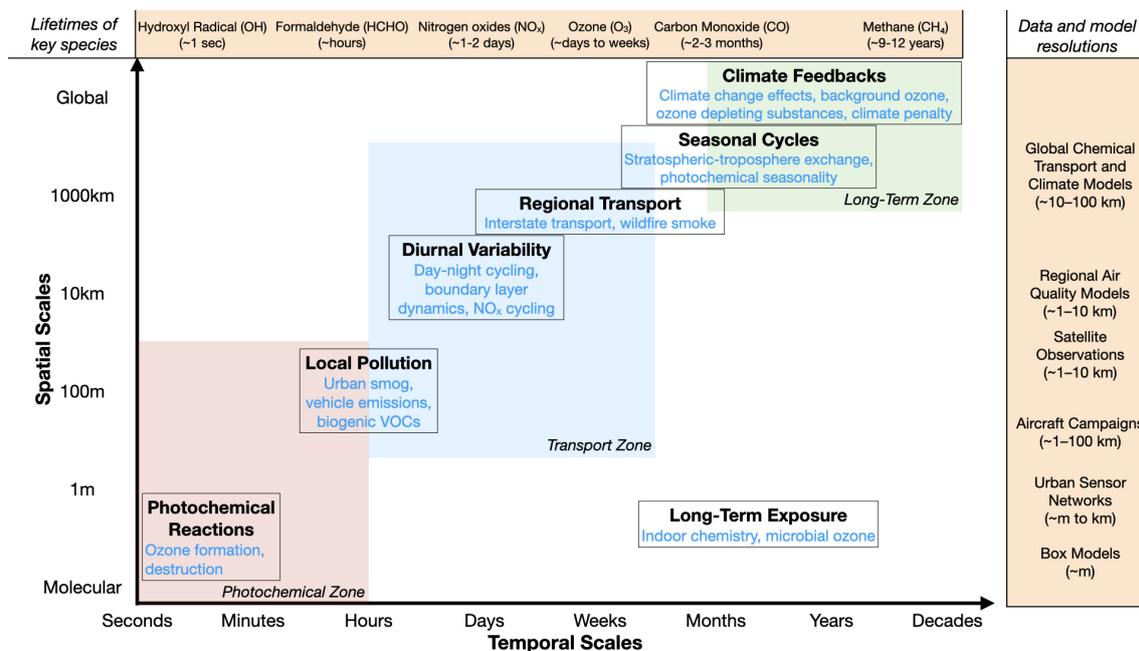
Tropospheric ozone is a harmful atmospheric pollutant and an important greenhouse gas, contributing to both environmental and public health issues. Long-term exposure to elevated ozone levels is linked to hundreds of thousands of premature deaths globally each year (Malashock et al., 2022; Malley et 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 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<sub>4</sub>), volatile organic compounds (VOCs), and nitrogen oxides NO<sub>x</sub> (NO<sub>x</sub>, NO+NO<sub>2</sub>). 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).

Modeling tropospheric ozone is therefore a challenging task due to the complex coupling of these chemical and physical processes that control its local concentrations across different spatial and temporal scales, as detailed in Figure 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, there are still challenges to reach accurate simulations of ozone (Young et al., 2018). For example, although ozone is the longest- and most-measured trace gas in the observational record, large uncertainties still exist in global model estimates of tropospheric ozone and its trends. 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<sub>x</sub>-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 Figure 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, Maryle-

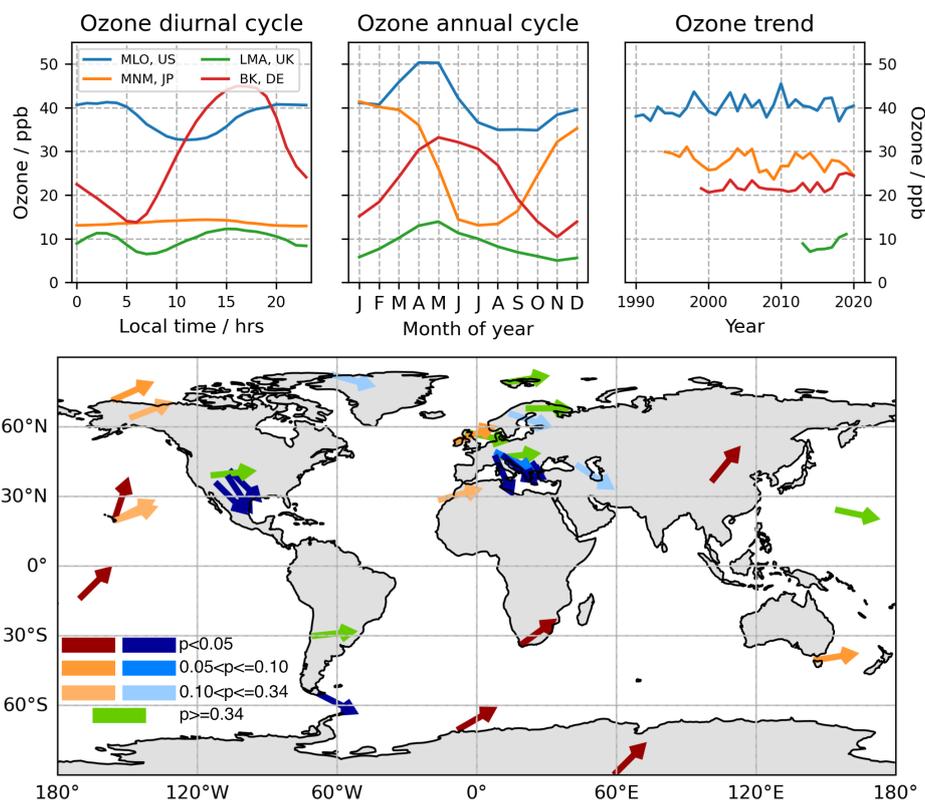


**Figure 1.** Spatial and temporal scales of tropospheric ozone chemistry processes. The x-axis shows timescales, from rapid photochemical reactions to long-term climate feedbacks, and the y-axis shows spatial scales, from local pollution to global atmospheric transport. Species lifetimes and relevant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

bone 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 long-term trends in ozone, although weak, also vary between the sites, with both modest increases (London Marylebone Road) and decreases (Minamitorishima) being seen. The lower panel shows variation in ozone trends across the globe, ranging between -3 and 3 ppbv per year, across remote sites.

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.

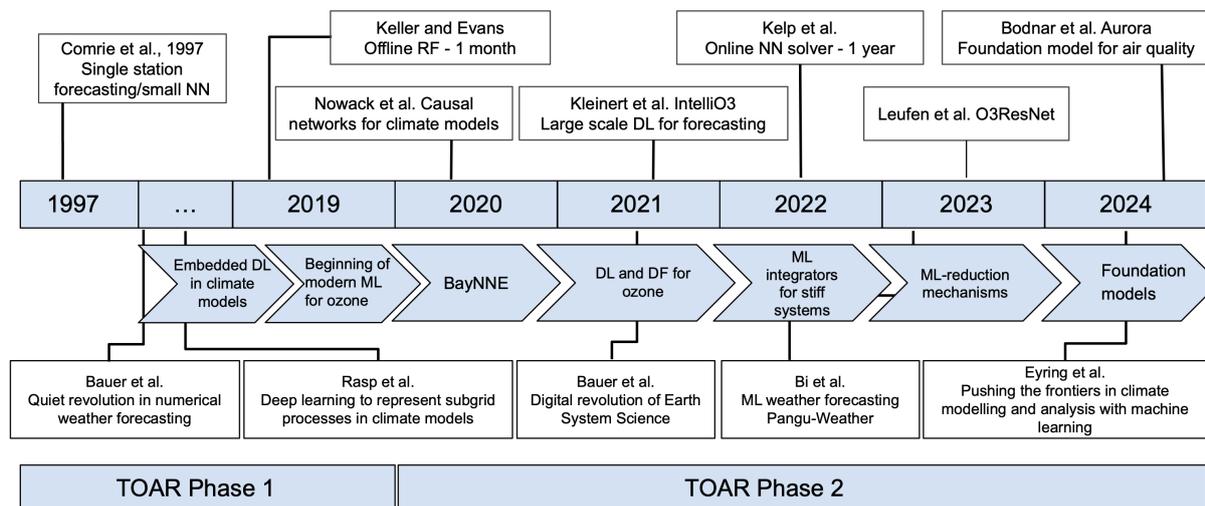
Figure 3 highlights progress in the field of ML as applied to weather and climate science, which has been rapid since the publication of the first phase of the TOAR assessment. In their review of the state of the field of weather forecasting, Bauer et al. (2015) note many areas of progress for the field, 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



**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. Lower panel: Long-term ozone trends based on monthly anomalies at remote surface sites. Data from Cooper et al. (2020) and replotted here

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).

While most ozone ML studies to date have prioritized prediction, ML also has great potential to enhance process-level understanding. Observational data, when integrated with model simulations through data assimilation techniques, have already improved the understanding of emissions and atmospheric chemistry by reducing uncertainties (Miyazaki et al., 2020b). 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



**Figure 3.** Timeline of a selection of studies using ML in ozone research, aligned with a selection of papers using ML in wider weather and climate modeling research. 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.

75 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 detail the bottlenecks to progress in applying ML to ozone problems, including the limitations of available training data, heterogeneity of data sources, and computational limits. We suggest that the use of ML can provide tangible improvement over existing methods in some aspects of ozone research and we provide possible paths towards further improvement to complement existing models and initiatives.

80 We highlight three areas where ML for ozone has been most widely applied: forecasts based on ground-based observations are reviewed in Section 2, methods for complementing or replacing parameterizations in numerical models of atmospheric chemistry and transport are discussed in Section 3, and ML models that are using satellite data or combined data products are presented in section 4. Limitations specific to each section are briefly mentioned, and we find common limitations across domains. Section 5 highlights and further details these cross-cutting issues and limitations with the application of ML to ozone studies, while Section 6 describes future directions for the field, highlighting emerging approaches that seek to address the cross-cutting challenges.



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

### 90 2.1 Background

State-of-the-art air quality forecasts, typically on the timescales of hours or days and up to a few days ahead, are based on the output of numerical chemical transport models (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  
95  $PM_{2.5}$  (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 Figure 1.

100 The short-term forecasting of air pollutants including ozone, i.e. predictions of expected concentrations over 1-4 days, is relevant for public health and scientific questions. One important task is to provide the public with accurate and timely warnings of poor air quality, which may influence public behavior and encourage precautions, leading to reduced exposure to poor air quality (Buonocore et al., 2021; Hahm and Yoon, 2021; Alari et al., 2021; Saberian et al., 2017). Existing ML forecast methods are well suited to this task. Similar ML architectures and workflows have also been used to fill in missing data in  
105 observational datasets (Betancourt et al., 2022; Wu et al., 2024).

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 chemical transport  
110 model (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 locations 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,  
115 increasing quantities of data and second, larger models with more appropriate inductive biases. The field has a long history (see Figure 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 neural network (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  
120 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  
125 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  
130 measurements. The introduction of the Tropospheric Ozone Assessment Report (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 curation of large datasets for scientific progress in ML is highlighted below in the Outlook section.

Increasingly, more complex architectures are being used to enhance the accuracy of ozone forecasts, and more data are  
135 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. In weather studies there is work on forecasting at observation stations using these methods, and  
140 transferring these methods to forecast ozone at ground-level stations is feasible (Manshausen et al., 2024).

General trends have emerged from recent forecasting studies with ML. Typically, models trained on many sites perform better than models trained to predict ozone at a single station, and larger deeper models are better able to forecast given sufficient data. However, it is difficult to determine precisely which individual features of models are contributing most to improved accuracy as studies typically use different test data. The establishment of a robust benchmark dataset, which could be  
145 used across the community to evaluate the performance of different methodologies would allow rigorous comparison between competing methods. A dataset of this kind for weather forecasting, WeatherBench, and its successor (Rasp et al., 2020, 2024) continue to facilitate effective inter-comparison between different ML and numerical models. 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. This benchmark dataset would need to cover a range of environments, including extreme  
150 events. 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).

### 2.3 Future Outlook

While forecasting ground-level ozone has seen some successes, there are still some necessary steps before we have reliable operational ML forecasts. As in other fields, the community has seen improved performance with scaling datasets, and training  
155 across a variety of environments with rich data (Leufen et al., 2023). While the TOAR database of surface observations provides



a large collection of in-situ data, the community is still missing a fused, global and gridded benchmark dataset that could be used to robustly compare methods (Dueben et al., 2022; Rasp et al., 2024), and explore model deficiencies. Creating this dataset should be a collaborative effort that may provide the step-change in ground level air pollution forecasting that other benchmark datasets have precipitated in other fields (Deng et al., 2009). Part of this benchmark should focus on extreme events and probabilistic metrics, as currently the community does not evaluate performance on extremes thoroughly, and encouraging probabilistic models and metrics may improve performance on extremes. While generalization is less of an issue for short-term forecasting, building models that can generalize to undersampled regions of the world would facilitate better pollution forecasting in countries with fewer data. Finally, work on interpretability and explainability is limited, and further study may give more insight into the internal working of these models, and possibly insight into physical processes.

### 165 3 Applications of ML methods in atmospheric chemical 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 chemistry in these models is typically represented by chemical transport models (CTMs), which focus on the distribution and chemical evolution of species in the atmosphere. For some applications, chemistry-climate models (CCMs) may further 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 or 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).

180 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 global/regional atmospheric models and ESMs typically include  $\mathcal{O}(100)$  coupled species and is a multiscale system 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  
190 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  
195 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). Additionally, while ML models are commonly trained and deployed using  
software libraries written in Python, the successful integration of ML-based algorithms and parameterizations within 3-D CTMs  
remains limited. A potential barrier is the difficulty of calling Python libraries from Fortran-based models, which Fortran cannot  
do so natively. However, recent advances in ML software can package ML code into static or dynamic libraries for flexible  
200 deployment.

There are a variety of examples using output from CTM and air quality models in tandem with satellite measurements,  
ground-level monitoring observations, meteorological reanalysis data, and emissions databases to predict or improve forecasts  
of ozone. These improvements can be achieved through various methods, including data fusion techniques and the creation of  
reduced-order models. However, because such studies involve additional data besides model outputs, they will be discussed in  
205 other sections (Sections 2 and 4). Here, we emphasize studies primarily employing data from atmospheric chemical models.

## 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.  
210 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) and offer insight  
beyond that of traditional model-to-observation comparison methods.

### 3.2.1 ML emulation and reduced order modeling

215 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-day forecast periods. Kuo and Fu (2023) investigated how accurately  
ML models can learn the ozone-NO<sub>x</sub>-VOC chemical relationships in a chemical mechanism and found that their ML model  
produced distorted NO<sub>x</sub> and VOC-limited isopleths when only trained on CMAQ model outputs. Kelp et al. (2022) trained an  
NN integrator in a photochemical box model, including an encoder/decoder to decrease dimensionality, and a recursive feed-  
220 back loop over 24-hr 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-day 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 ML models implemented within global CTMs

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 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 an average NMB of 2.63% 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 days and stability over longer time scales (>1 year) remains to be seen.

Few studies have implemented ML models directly within CTMs/ESMs to evaluate their effectiveness, although the necessary tools exist. 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. See review by de Burgh-Day and Leeuwenburg (2023) for greater discussion on using ML for facilitating speedups and computational cost reductions in atmospheric science models.



### 255 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. Wisner 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  
260 gradient-boosted regression trees to develop a parametrization for the OH radical, a key driver of O<sub>3</sub> 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.

Silva et al. (2019) developed an ML parameterization for ozone dry deposition velocities using surface observations that  
265 outperformed those within CTMs for certain locations. Similarly, Ivatt and Evans (2020) created an 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. To be sure, there is a growing literature on ML approaches for bias corrections on existing air quality modeling systems (Neal et al., 2014; Borrego et al.,  
270 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. 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 random forests 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 a random forest model to  
275 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 tracers within  
280 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  
285 emulation for atmospheric modeling (Krasnopolsky et al., 2005, 2008).



### 3.3 Future Outlook and Priorities

The near-term future of integrating ML with 3-D atmospheric chemistry and climate modeling, particularly for tropospheric ozone prediction, 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. Additional priorities should include scaling regional ML models to a global context and identifying the strengths of ML (e.g., speed of operation, pattern recognition of high volumes of data) alongside its weaknesses (e.g., inability to approximate multiscale dynamics, instability over time, resource-intensive training and black box nature).

In the context of using ML for model component emulation and replacing traditional components, a major challenge is managing coupled instabilities, where ML predictions accumulate errors and drift from the reference model over time. This instability is also prevalent in ML-based model emulation for weather and climate applications. Consequently, 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 and 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.

Another emerging area is the use of Physics-Informed Neural Networks (PINNs), which embed physical laws into an ML model's architecture. While ML has been used to emulate traditional numerical solvers, this approach has often sacrificed accuracy and stability over long-term timescales in exchange for faster computations. Traditional solvers, though computationally expensive, do not exhibit these error instabilities (see Le Chatelier's principle). PINNs may offer a solution to balance speed with accuracy and have shown success in smaller-scale systems, such as single ODEs or PDEs (Karniadakis et al., 2021). Although their application to large groups of ODEs in atmospheric chemistry has not yet been widely explored, PINNs may hold the potential for achieving greater accuracy than traditional solvers, possibly overcoming some of the limitations of ML approaches in atmospheric chemistry modeling. However, PINNs may not enforce physical laws perfectly and the inherent black-box nature of certain ML components limits their transparency compared to physical models, which provide clearer constraints. Therefore, while PINNs may address some computational challenges, their contribution to advancing process-level understanding in atmospheric chemistry remains uncertain.

At present, most ML models attempting to replace components of CTM/CCM parameterizations have used supervised learning methods. These methods typically train using chemical data from box models or other solvers, aiming to faithfully reproduce outputs given initial conditions. In contrast, unsupervised learning, which identifies patterns or clusters in data without explicit labels, remains underexplored in this field. Greater emphasis is needed on understanding the factors influencing ML model performance, whether stemming from the ML model architecture, the quality and curation of training data, or uncertainties inherent in the physical processes of emissions, transport, and chemistry. To be sure, the advent of transformers



320 and Large Language Models (LLMs) offers a new frontier for ingesting and potentially improving AI modeling of atmospheric chemistry, but it will be crucial to understand the above points before pursuing even more complicated ML model architectures.

## 4 Applications of AI/ML methods to satellite observations

### 4.1 Background

325 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. Particularly in remote areas where in-situ observations are not available, they can fill the gaps in the sparse distribution.

Over the past few decades, multiple satellites have been launched to measure total ozone columns, including the Total Ozone Mapping Spectrometer (TOMS) (Prospero et al., 2002), Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) (Bovensmann et al., 1999), Ozone Monitoring Instrument (OMI) (Veefkind et al., 2006), Global Ozone Monitoring Experiment (GOME) (Coldewey-Egbers et al., 2005), GOME-2 (Loyola et al., 2011) and the Sentinel-5P Tropospheric Monitoring Instrument (TROPOMI) (Garane et al., 2019). Geostationary satellite observations, such as the Tropospheric Emissions: Monitoring of Pollution (TEMPO) (Naeger, 2021) and Geostationary Environment Spectrometer (GEMS) (Baek et al., 2023) observe detailed spatial and temporal patterns of total column ozone and other air pollutants. 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 TES (Bowman et al., 2002) and IASI (Boynard et al., 2009), and by combining measurements from both ultraviolet (UV) and visible (VIS) wavelengths by TEMPO (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 IAGI+GOME-2 (Cuesta et al., 2018) and TES+OMI (Colombi et al., 2021), have been implemented to derive tropospheric ozone profiles with increased sensitivity to the lower troposphere.

345 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<sub>2</sub> and CH<sub>2</sub>O from OMI, GOME-2, TROPOMI, and OMPS, 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. AI/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.



## 350 4.2 Progress and State of the Science

AI/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. AI/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.

### 355 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. Usually carried out with numerical calculations based on physics, using ML for ozone retrieval is a burgeoning field. The task is to take level 1 radiance data from the satellites and to produce a level 2 (L2) product, called retrievals, such as vertical profiles and vertical columns.

360 AI/ML-driven retrieval algorithms have emerged as a powerful alternative to improve the processing efficiency of atmospheric composition satellite products. Physics-based retrievals, which are based on radiative transfer models (RTMs) and solves their inverse problem, have been widely used to generate satellite L2 products of atmospheric composition concentrations 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 O<sub>3</sub> (e.g., Müller et al., 2003), SO<sub>2</sub> (e.g., (Li et al., 2022)), isoprene (e.g., (Wells et al., 2022)), and CO<sub>2</sub> (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 GEMS operational L2 NO<sub>2</sub> 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+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, 375 emission 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 satellite. Satellite observations of ozone and its precursors, combined with additional information such as meteorological conditions, land-use, 380 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). They contribute to air quality monitoring and human health impact assessment, with some studies focusing particularly on estimating health-relevant metrics using remote sensing data.

385 Di et al. (2017) proposed a hybrid neural network model using satellite-based data from OMI, GEOS-Chem CTM outputs, ozone vertical profiles, meteorological variables, land-use terms and other atmospheric compounds to predict MDA8 ozone in the continental United States. eXtreme Gradient Boosting (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 1km resolution in Taiwan. Ghahremanloo et al. (2023) used a CNN with TROPOMI data as an input to estimate MDA8 in the United States.  
390 Among various ML techniques, Zong et al. (2024) concluded that Deep Forest performs better than other shallower, 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<sub>2</sub> products to estimate surface NO<sub>2</sub> concentrations with high spatial resolution ((Kim et al., 2021), and to satellite aerosol optical depth measurements to estimate surface PM<sub>2.5</sub> concentrations (Huang et al., 2021; Xiao et al., 2021) which are useful for exposure estimates.

395 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., 2024; Xing et al., 2022; Tu et al., 2023; Li et al., 2024; Bruno et al., 2024).

In addition, AI-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 neural network to identify and quantify point source emissions of methane from hyperspectral images from the PRecursor IperSpettrale della Missione Applicativa (PRISMA) satellite with 400 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 the hydroxyl radical (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). Zhu et al. (2022) developed a ML approach as an efficient alternative to computationally expensive CTM-based OH simulations, using observations of tropospheric NO<sub>2</sub> and HCHO columns from OMI as input to a ML model to predict daily near-surface 405 OH. Anderson et al. (2023) used a combination of ML models, CTM simulations, and various satellite observations to estimate tropospheric column OH in the tropics over the open oceans. They also showed that systematic biases in satellite products such as NO<sub>2</sub> may lead to biases in OH estimation.

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 415



the characteristics and quality of satellite products, selection of effective supplementary information, and further development of appropriate ML methods.

### 4.3 Future Outlook and Priorities

Satellite observations are available at regional and global scales, for temporal scales ranging from hourly to decadal scales. As with the integration of ML with CTMs, priorities in the area of ML using satellite observations need to be further focused on their appropriate application at various spatio-temporal scales. With increasing resolution of satellite observations, the generalization of methods for producing satellite retrievals by ML has also become one of the priorities of the community.

Each satellite measurements have different characteristics, and data quality may vary not only spatially but also temporally. In many areas, especially in developing countries where air pollution is severe, ground truth observations are not available, hindering the application and validation of ML models. Careful validation of ML results is desirable, for example, taking into account the differences in spatial representativeness between satellite and ground-based observations. Another important challenge is how to integrate and use the results of validation using independent observations and uncertainty information from satellite retrievals as input for ML.

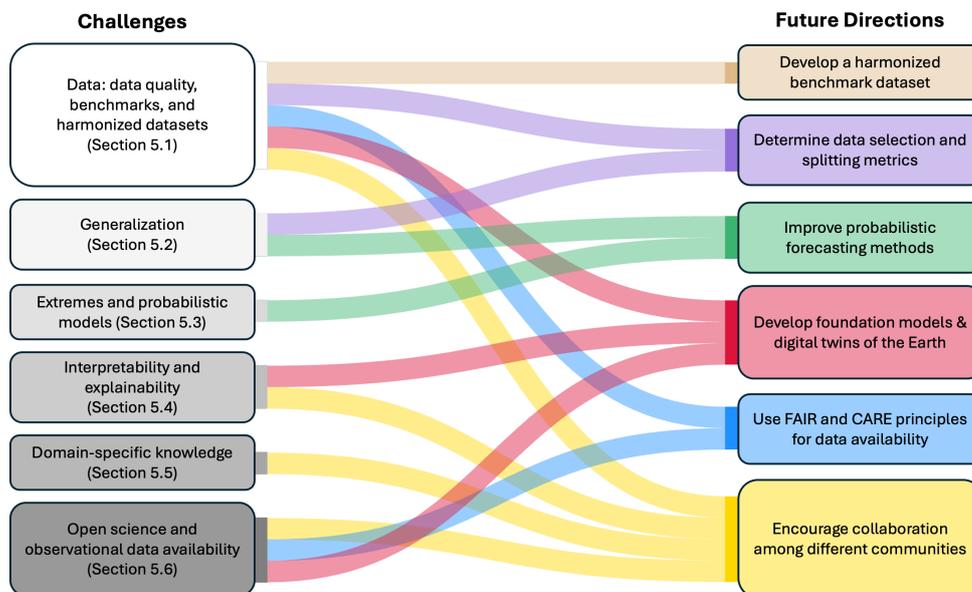
As the satellite constellation continue to evolve, data fusion of multiple instruments and multiple species using ML, in addition to chemical reanalysis approaches using data assimilation techniques (Inness et al., 2019; Miyazaki et al., 2020a), will play an important role in achieving consistent analysis of long-term trends. Combining satellite data with ML offers a further approach to investigate the linear and non-linear processes controlling air pollutant dynamics, complementing ground-based measurements and CTM estimates.

## 5 Challenges and Limitations

In this section, we reflect on some common challenges and limitations of using AI in the context of ozone forecasting, modeling and observations (Figure 4).

### 5.1 The challenges of data availability and workflow

Datasets are critically important to ML/AI research, and along with model choice and training methods, i.e. workflow, central to the success of modeling efforts and to their utility. 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 a key factor driving the transformation of ML weather forecasting between 2022 and 2024. Regarding surface ozone data, differences in the spatial coverage of ozone or of its precursors data will affect model development and the performance of data-hungry ML models. Ultimately, a lack of sufficient surface observations will impact the study of air quality and downstream impacts on health or vegetation. Therefore



**Figure 4.** Challenges and future directions described in Sections 5-6. The left hand side represents the categories of challenges described in further detail in the sections listed. The right hand side lists future directions and priorities for AI/ML ozone studies. The lines connect the challenges with considerations for future work on the right hand side.

improving data coverage over poorly monitored areas remains a priority (Schultz et al., 2017). In contrast, the breadth of information available in datasets like TOAR, GHOST and reanalysis products like CAMS and TC 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 AI 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 section 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 model training, there is a wide array of data-splitting approaches that can answer subtly related scientific questions. As the motivating prediction challenges use both structured data and have structured prediction tasks, it is important



to match the way data are split to the scientific task at hand. 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. 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 and some solutions have been developed in other ML applications. However, 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.

## 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. Many ML models, particularly those trained as supervised learning problems, act as high-dimensional statistical interpolators. Consequently, these models may not genuinely learn the physics and chemistry of the atmospheric system but instead, learn a mapping from inputs to outputs based on the bounds of the training data. One significant challenge faced by these ML models, whether in supervised or unsupervised settings, is concept drift. Concept drift refers to ML predictions becoming less accurate over time, which can arise from a non-representative training dataset (initial conditions are narrow, do not sample enough timescales) 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 when fine-tuned from previously well-trained models, while also demonstrating improved accuracy and stability over time.

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. For example, if a model is trained on data from one country with specific emissions and climate, it may not be expected that this model will perform as accurately in another country with a very different distribution of data. 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 generalise to unseen conditions and perturbations (Hakim and Masanam, 2024). Generalisation 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, using models that report uncertainty in unfamiliar domains.

To what extent these results would apply to large-scale ML models for ozone remains to be seen. For tropospheric ozone, the correctness of capturing atmospheric chemical processes that are not always directly observable, but which can have an influence on ozone concentrations, particularly on longer time scales, may be important. Examples of this include HONO, nitrate and aldehyde formation and conversion processes, which provide for “storage of radicals” during nighttime. These radicals can be released with sunrise in the morning hours and drastically increase ozone formation rate. While it may be possible that sufficiently complex ML models capture enough auxiliary information to model this ozone boost correctly, such case studies will become important and have not yet been explored.

### 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 (LLHI) 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. Naturally, extreme events are rarely and infrequently represented in data, and therefore supervised ML approaches trained with non-extreme data may fail to predict extremes well. 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 (Bodnar et al., 2024). Well calibrated forecast uncertainty is important for delivering accurate air quality warnings and for public information. Furthermore for other tasks where data quality is critical, such as infilling missing data for use in later analysis, well-calibrated estimates of uncertainty provide information on the quality of ML generated data and hence on further analysis. This is especially relevant as many infilling approaches use average values, and therefore may not be applicable for extremes.

#### 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 the internal data transformations are comprehensible to us, 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 neural networks 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 neural networks more generally have been studied widely, mechanistic interpretability of neural networks 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.



## 5.5 The challenges arising from domain-specific knowledge

565 Modeling ozone using ML proves challenging due to the multitude of sources driving model error (emissions, chemistry, trans-  
port, deposition, representativeness) 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 adap-  
570 tive 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 atmo-  
spheric chemistry can help identify the optimal training dataset and define meaningful loss functions and targeted timescales  
(Figure 1) for the ML model problem. 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). Knowledge  
of slow chemical processes, such as the role of peroxyacetyl nitrate (PAN) decomposition for ozone formation over polluted  
575 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.

Alternatively, 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  $\text{NO}_x$ -limited vs. VOC-limited  
580 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, while useful for capturing dominant signals, risks overlooking patterns  
that could be hidden in the chemical state space that may promote greater accuracy and stability over longer time scales. While  
these uncovered relationships may not always represent true causal mechanisms, the value of ML lies in its capacity to dis-  
585 cover interactions that may escape human intuition. 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

To thrive, the interdisciplinary ozone modeling and forecasting community requires open knowledge sharing, open resources  
and research cooperation. Research in the domain should adhere to the FAIR principles of Findability, Accessibility, Interop-  
590 erability and Reusability (Wilkinson et al., 2008) and the CARE principles of Collective, Authority to control, Responsibility  
and Ethics (Stephanie Russo Carroll et al., 2021). Availability of data is essential for data-driven approaches and the developed  
TOAR-II surface ozone database is key here through its open data policies and its Application Programming Interface (API),  
which allow for automatic extraction of data and integration into a workflow (e.g. in Jupyter Notebooks). Besides the afore-  
mentioned "FAIR data", a call for FAIR-software has been recognized and expressed more recently (Barker et al., 2022), thus  
595 including a larger part of the research cycle than only data collection, storage and dissemination.



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). The increasing resolution of satellite products and models is often considered to be an improvement, but the larger data size can  
600 complicate the processing and analysis of data for some researchers (Jain et al., 2022). The concept of data availability may thus be perceived differently across regions. There are also additional barriers accessing data. 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  
605 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 relevant, global data relevant for ozone modeling studies, ranging from land-use information from MODIS (Friedl, 2021) to human modification data from  
610 VIIRS nighttime lights (Elvidge et al., 2017), Gridded Population of the World (University, 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.

## 615 **6 Future Directions**

ML advances are swiftly revolutionizing atmospheric composition research in general, and ozone thermodynamics and chemical processing in particular, harnessing both Earth observations and modeling outputs to extract valuable insights from vast datasets.

A wealth of Earth observation data from satellites, in-situ measurements, and ground-based instruments is now available,  
620 with data volumes already well beyond tens of petabytes. This abundance of data is marking a golden age for Earth observation. New developments in deep learning (DL) are rapidly overcoming limitations related to spatio-temporal and multi-variable relationships, promising significant breakthroughs (Eyring et al., 2024). However, their direct application to Earth observations is often hindered by challenges such as data quality and coverage, model generalization across different regions, and the complexity of associated metadata, such as relative position, orientation, and shading in satellite imagery. Future research and  
625 advancements in ML-based observational products, including efforts to address uncertainty quantification e.g. (Haynes et al., 2023), will further enhance our understanding, facilitate process-based model evaluation (Nowack et al., 2020), and enable actionable science, particularly in areas such as human health (Fleming et al., 2018) and climate impact (Gaudel et al., 2018) assessments.



Accurate forecasts of extrema in short-term surface O<sub>3</sub> predictions are essential for protecting human health, while reliable  
630 projections of long-term changes in tropospheric O<sub>3</sub> abundances are critical for understanding climate change and avoiding its  
most severe 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). Combining causal discovery and eXplainable Artificial  
Intelligence (XAI) methods holds potential for advanced process-based evaluation (Iglesias-Suarez et al., 2024). This approach  
can enable the identification of physical and chemical processes that contribute to both good and poor model performance,  
635 providing valuable insights into the drivers of ozone, including sources and sinks. On the other hand, these methods are  
generally limited to small-scale ML models and different approaches (e.g. analysis of attention maps) need to be sought for  
large-scale ML models. 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 and validation of ML-based ozone modeling techniques. Yet, current domain-specific ML methods and tailored  
640 solutions often lack a more universal perspective, whereas general-purpose ML-based algorithms would only need to be fine-  
tuned to address specific tasks.

Nevertheless, within the context of ML-based solutions for atmospheric composition, including ozone, and numerical mod-  
eling, future work can focus on foundation models to advance more integrated approaches. Foundation models, as large-scale,  
pre-trained architectures, are emerging as promising tools for atmospheric modeling. These models, built on extensive datasets  
645 in a self-supervised manner Bommasani et al. (2021), have already demonstrated their adaptability in fields like weather fore-  
casting and climate science (Lessig et al., 2023; Nguyen et al., 2023; Bodnar et al., 2024). In the context of tropospheric  
ozone modeling, foundation models could reduce computational burdens while learning from varied datasets, including obser-  
vational 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  
650 potentially robust air quality benchmarks by supplementing and harmonizing observational data. Their flexible architecture  
enables training a single model with large-scale resources and fine-tuning it for multiple tasks, reducing the computational ex-  
pense of repeated model training (Bommasani et al., 2021). For ozone forecasting, they offer a path toward improved modeling  
capabilities by leveraging diverse data sources and capturing complex relationships in atmospheric composition. Recent work  
has shown that foundation model architectures can simulate atmospheric dynamics and composition by emulating variables  
655 from reanalysis products like CAMS (Inness et al., 2019), achieving promising results for multi-day air pollutants forecasting,  
including ozone (Bodnar et al., 2024). This shift would, in turn, facilitate the development of models that are not only specific  
to ozone but capable of generalizing to broader air quality applications, potentially leading to more robust and versatile ML  
solutions.

For CTMs, particularly, and ESMs more broadly, foundation models can serve as valuable tools to either complement or  
660 enhance traditional approaches. Chemistry schemes and parameterizations based on ML models can bring computationally  
efficient interactive chemistry to ESMs for accurate long-term projections (Eyring et al., 2016). Additionally, end-to-end em-  
ulators can complement physics-based simulations by employing large ensembles to better constrain uncertainties. While  
general-purpose ML algorithms, such as foundation models, can provide a step change in data-driven approaches, they may



not always lead to actionable science, such as contributing to more informed decision-making by, for example, accurately  
665 attributing ozone pollution events to specific sources— and the development of effective mitigation strategies.

The need for actionable science to address challenges like climate change and air quality has led to the development of  
digital twins of the Earth (DTEs), which integrate ML models and physical simulations to provide comprehensive monitoring  
and forecasting of the Earth system (Bauer et al., 2021; Bauer, 2024). DTEs enable users to explore “what-if” scenarios and  
develop targeted strategies, such as mitigating ozone pollution and improving air quality, by simulating the impacts of various  
670 interventions. This user-centric approach contrasts with traditional top-down methods and supports more effective, adaptive  
responses to environmental challenges. However, the implementation of DTEs raises critical issues regarding governance,  
data accessibility, and computational demands, particularly as they rely on exascale computing to deliver real-time insights  
(Hazeleger et al., 2024). As global initiatives like the European Commission’s Destination Earth and NASA’s Earth System  
Digital Twin advance these frameworks, ML techniques will be essential for ensuring their efficiency and scalability, enabling  
675 DTEs to serve as a transformative tool for improving air quality and addressing broader environmental challenges.

To meet society’s needs facing current environmental challenges by providing actionable science and maintaining the rapid  
progress in this field, collaboration among different communities is essential. The full potential of ML in atmospheric compo-  
sition in particular, and Earth system science in general, can only be realized through an interdisciplinary approach, fostering  
close cooperation between domain-knowledge and ML communities. Furthermore, this requires opening new collaborative  
680 opportunities between academia and the private sector. As awareness grows within the ML community regarding the societal  
relevance of algorithms in Earth system research, major technology firms are increasingly interested in applying their expertise  
to environmental-related issues. They may well pursue this through interdisciplinary research collaborations with domain-  
expert scientists, whether through direct employment or academic partnerships. As these applications extend into unexplored  
future climates, input from academic experts will become increasingly crucial, facilitating collaborative efforts to advance the  
685 understanding of the Earth system.

## 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, chemical mechanism emulation and remote sensing of ozone. Specifically, ML methods are providing increas-  
690 ingly 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.

In addition to excelling at specific tasks, recent work illustrates that foundation models, trained on diverse datasets, are  
capable of more general atmospheric composition modeling, including ozone. This paradigm of foundation models represents  
695 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. For this work to make further progress in modeling real-



world ozone faithfully, models should be trained with a synthesis of high-quality observational datasets. In order to facilitate robust, rapid progress in ML for ozone modeling, 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 challenges faced by existing ML models is necessary, which will require close collaboration between domain experts and ML researchers to develop models tailored to the specific challenges of ozone modeling. Finally, it remains an open and important question whether ML models can contribute to improved process-level understanding of the drivers of ozone, and generalise 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 meaningful 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

### 710 **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.

### 715 **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):**

720 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.

### 725 **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.

730 **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.

735 **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.

740 **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.

745 **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.

750 **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):**

755 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.

### 760 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.

### 765 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 (ML4O3, 2024).

*Author contributions.* All authors contributed to the writing and review of the manuscript. SHMH, MK, PTG, F-IS, GK, KD, KZ, MGS and EAP led the writing and preparation of the MS. SHMH led Section 2, MK section 3, KD, EAP and KM led section 4, GK section 5 and FI-S  
770 section 6.

*Competing interests.* S.H.M.H., M.K., D.E.C., P.T.G. declare no competing interests. Some of the authors are involved with editorial work for Copernicus journals.

*Acknowledgements.* S.H.M.H. acknowledges funding from EPSRC via the AI4ER CDT at the University of Cambridge (EP/S022961/1). P.T.G. acknowledges the Environmental Geochemical Cycle Research Group, Earth Surface System Research Center, Japan Agency For  
775 Marine Science and Technology for support as an External Researcher. P.T.G. and A.T.A. 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). Z.L. was acknowledges the National Natural Science  
780 Foundation of China, grant number 42307140. MGS acknowledges funding by the European Commission under grant ERC-AdvG-787576. K.-L.C. was supported by NOAA cooperative agreement (no. NA22OAR4320151). J.J.W. acknowledges National Aeronautics and Space Administration (NASA) grants no. NNX16AQ30G and no. 80NSSC23K0930. D.E.C. acknowledges financial support from Underwriters Laboratories, Inc.



## References

- 785 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>, publisher: Taylor & Francis tex.eprint: <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, 106 583, 2021.
- 790 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, *Geoscientific Model Development*, 15, 6341–6358, <https://doi.org/10.5194/gmd-15-6341-2022>, 2022.
- Anderson, D. C., Duncan, B. N., Nicely, J. M., Liu, J., Strode, S. A., and Follette-Cook, M. B.: Technical note: Constraining the hydroxyl (OH) radical in the tropics with satellite observations of its drivers – first steps toward assessing the feasibility of a global observation strategy, *Atmospheric Chemistry and Physics*, 23, 6319–6338, <https://doi.org/10.5194/acp-23-6319-2023>, publisher: Copernicus GmbH, 2023.
- 795 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.
- 800 Arroyo, A., Herrero, A., Tricio, V., Corchado, E., and Wozniak, M.: Neural models for imputation of missing ozone data in air-quality datasets, *Complexity*, p. 7238015, <https://doi.org/10.1155/2018/7238015>, 2018.
- 805 Baek, K., Kim, J. H., Bak, J., Haffner, D. P., Kang, M., and Hong, H.: Evaluation of total ozone measurements from Geostationary Environmental Monitoring Spectrometer (GEMS), *Atmospheric Measurement Techniques*, 16, 5461–5478, <https://doi.org/10.5194/amt-16-5461-2023>, publisher: Copernicus GmbH, 2023.
- Balagus, 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, *Atmospheric Measurement Techniques*, 16, 3787–3807, <https://doi.org/10.5194/amt-16-3787-2023>, publisher: Copernicus GmbH, 2023.
- 810 Barker, M., Chue Hong, N., and Katz, D. e. a.: Introducing the FAIR principles for research software, *Scientific data*, 9, 622, <https://doi.org/10.1038/s41597-022-01710-x>, 2022.
- Bauer, P.: What if? Numerical weather prediction at the crossroads, *arXiv preprint*, 2024.
- 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.
- 815 Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., and Wedi, N. P.: The digital revolution of Earth-system science, *Nature Computational Science*, 1, 104–113, <https://doi.org/10.1038/s43588-021-00023-0>, publisher: Nature Publishing Group, 2021.
- 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, <https://doi.org/10.21203/rs.3.rs-5356949/v1>, iSSN: 2693-5015, 2024.



- 820 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., and Stadtler, S.: AQ-Bench: a benchmark dataset for machine learning on global air quality metrics, *Earth System Science Data*, 13, 3013–3033, 2021.
- Betancourt, C., Stomberg, T. T., Edrich, A.-K., Patnala, A., Schultz, M. G., Roscher, R., Kowalski, J., and Stadtler, S.: Global, high-resolution  
825 mapping of tropospheric ozone & explainable machine learning and impact of uncertainties, *Geoscientific Model Development Discussions*, pp. 1–36, <https://doi.org/10.5194/gmd-2022-2>, publisher: Copernicus GmbH, 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, 2024.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks,  
830 *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.:  
835 *Aurora: a foundation model of the atmosphere*, arXiv preprint 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 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.
- 840 Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIAMACHY: Mission Objectives and Measurement Modes, *Journal of the Atmospheric Sciences*, 56, 127–150, [https://doi.org/10.1175/1520-0469\(1999\)056<0127:SMOAMM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2), 1999.
- 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>,  
845 eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2002JD002150>, 2002.
- Boynard, A., Clerbaux, C., Coheur, P.-F., Hurtmans, D., Turquety, S., George, M., Hadji-Lazaro, J., Keim, C., and Meyer-Arneke, J.: Measurements of total and tropospheric ozone from IASI: comparison with correlative satellite, ground-based and ozonesonde observations, *Atmospheric Chemistry and Physics*, 9, 6255–6271, <https://doi.org/10.5194/acp-9-6255-2009>, publisher: Copernicus GmbH, 2009.
- Brasseur, G. P. and Jacob, D. J.: *Modeling of Atmospheric Chemistry*, Cambridge University Press, Cambridge,  
850 <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, *Atmospheric Measurement Techniques*, 17, 2625–2636, <https://doi.org/10.5194/amt-17-2625-2024>, publisher: Copernicus GmbH, 2024.
- Buonocore, J., Robinson, L., Hammitt, J., and O’Keeffe, L.: Estimating the potential health benefits of air quality warnings, *Risk Analysis*,  
855 41, 645–660, 2021.
- Casciari, G., Cavaola, M., and Mazzino, A.: Calibrating the CAMS European multi-model air quality forecasts for regional air pollution monitoring, *Atmospheric Environment*, 287, 119 259, 2022.



- Chaloulakou, A., Saisana, M., and Spyrellis, N.: Comparative assessment of neural networks and regression models for forecasting summer-time ozone in Athens, *Science of the Total Environment*, 313, 1–13, 2003.
- 860 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, 160 928, <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*, 865 112, 102 955, <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, e2021MS002 806, <https://doi.org/10.1029/2021MS002806>, [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021MS002806](https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021MS002806), 2022.
- 870 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?, preprint, *Gases/Field Measurements/Troposphere/Chemistry (chemical composition and reactions)*, <https://doi.org/10.5194/acp-2022-330>, 2022.
- 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., 875 Steiner, A. L., and Tai, A. P. K.: Dry Deposition of Ozone Over Land: Processes, Measurement, and Modeling, *Reviews of Geophysics*, 58, e2019RG000 670, <https://doi.org/10.1029/2019RG000670>, [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019RG000670](https://onlinelibrary.wiley.com/doi/pdf/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, ISBN: 1096-2247 Publisher: Taylor & 880 Francis, 2000.
- Coldewey-Egbers, M., Weber, M., Lamsal, L. N., de Beek, R., Buchwitz, M., and Burrows, J. P.: Total ozone retrieval from GOME UV spectral data using the weighting function DOAS approach, *Atmospheric Chemistry and Physics*, 5, 1015–1025, <https://doi.org/10.5194/acp-5-1015-2005>, publisher: Copernicus GmbH, 2005.
- Colombi, N., Miyazaki, K., Bowman, K. W., Neu, J. L., and Jacob, D. J.: A new methodology for inferring surface ozone from multi- 885 spectral satellite measurements, *Environmental Research Letters*, 16, 105 005, <https://doi.org/10.1088/1748-9326/ac243d>, publisher: IOP Publishing, 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?, *Atmospheric Chemistry and Physics*, 23, 4031–4044, <https://doi.org/10.5194/acp-23-4031-2023>, 2023.
- 890 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 895 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, *Atmospheric Chemistry and Physics*, 18, 9499–9525, <https://doi.org/10.5194/acp-18-9499-2018>, publisher: Copernicus GmbH, 2018.
- 900 Dadheech, N., He, T.-L., and Turner, A. J.: High-resolution greenhouse gas flux inversions using a machine learning surrogate model for atmospheric transport, *EGUsphere*, pp. 1–21, <https://doi.org/10.5194/egusphere-2024-2918>, publisher: Copernicus GmbH, 2024.
- de Burgh-Day, C. O. and Leeuwenburg, T.: Machine learning for numerical weather and climate modelling: a review, *Geoscientific Model Development*, 16, 6433–6477, <https://doi.org/10.5194/gmd-16-6433-2023>, 2023.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, in: 2009 IEEE  
905 conference on computer vision and pattern recognition, pp. 248–255, IEEE, <https://doi.org/10.1109/CVPR.2009.5206848>, 2009.
- 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>, publisher: Taylor & Francis \_eprint: <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  
910 and Recycling*, 184, 106 380, <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, 2022.
- Duncan, B., Anderson, D., Fiore, A., Joiner, J., Krotkov, N., Li, C., Millet, D., Nicely, J., Oman, L., St. Clair, J., Shutter, J., Souri, A., Strode, S., Weir, B., Wolfe, G., Worden, H., and Zhu, Q.: Opinion: Beyond Global Means: Novel Space-Based Approaches  
915 to Indirectly Constrain the Concentrations, Trends, and Variations of Tropospheric Hydroxyl Radical (OH), *EGUsphere*, pp. 1–35, <https://doi.org/10.5194/egusphere-2024-2331>, publisher: Copernicus GmbH, 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  
920 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, *Geoscientific Model Development*, 11, 2941–2953, <https://doi.org/https://doi.org/10.5194/gmd-11-2941-2018>, 2018.
- 925 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>, 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.
- 930 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., Bony, S., Meehl, G., Senior, C., Stevens, B., Stouffer, R., and Taylor, K.: Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958, 2016.
- 935 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*, pp. 1–13, 2024.
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., and Wang, J.: Artificial neural networks forecasting of PM<sub>2.5</sub> pollution using air mass trajectory based geographic model and wavelet transformation, *Atmospheric Environment*, 107, 118–128, 2015.
- 940 Finch, D. P., Palmer, P. I., and Zhang, T.: Automated detection of atmospheric NO<sub>2</sub> plumes from satellite data: a tool to help infer anthropogenic combustion emissions, *Atmospheric Measurement Techniques*, 15, 721–733, <https://doi.org/10.5194/amt-15-721-2022>, publisher: Copernicus GmbH, 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., Isak-  
945 sen, 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, D04 301, <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.,  
950 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, 025 008, <https://doi.org/10.1088/2752-5295/ac9cc2>, publisher: IOP Publishing, 2022.
- Fleming, Z., Doherty, R., Von Schneidemesser, E., Malley, C., Cooper, O., Pinto, J., Colette, A., Xu, X., Simpson, D., Schultz, M., and  
955 Lefohn, A.: Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health, *Elem Sci Anth*, 6, 12, 2018.
- Friedl, S.-M.: MCD12Q1 MODIS/terra+aqua land cover type yearly L3 global 500m SIN grid V006, <https://doi.org/10.5067/MODIS/MCD12Q1.006>, 2021.
- Garane, K., Koukouli, M.-E., Verhoelst, T., Lerot, C., Heue, K.-P., Fioletov, V., Balis, D., Bais, A., Bazureau, A., Dehn, A., Goutail, F.,  
960 Granville, J., Griffin, D., Hubert, D., Keppens, A., Lambert, J.-C., Loyola, D., McLinden, C., Pazmino, A., Pommereau, J.-P., Redondas, A., Romahn, F., Valks, P., Van Roozendaal, M., Xu, J., Zehner, C., Zerefos, C., and Zimmer, W.: TROPOMI/S5P total ozone column data: global ground-based validation and consistency with other satellite missions, *Atmospheric Measurement Techniques*, 12, 5263–5287, <https://doi.org/10.5194/amt-12-5263-2019>, publisher: Copernicus GmbH, 2019.
- Gardner, M. and Dorling, S.: Statistical surface ozone models: an improved methodology to account for non-linear behaviour, *Atmospheric Environment*, 34, 21–34, 2000.
- 965 Gaudel, A., Cooper, O., Ancellet, G., Barret, B., Boynard, A., Burrows, J., Clerbaux, C., Coheur, P., Cuesta, J., Cuevas, E., and Doniki, S.: Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation, *Elem Sci Anth*, 6, 39, 2018.
- 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, 121 508, <https://doi.org/10.1016/j.envpol.2023.121508>, 2023.



- 970 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, *Atmospheric Chemistry and Physics*, 24, 3163–3196, <https://doi.org/10.5194/acp-24-3163-2024>, 2024.
- 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., pp. 287–422, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/9781009157896.004>, section: 2 Type: Book Section, 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, 064 048, 2021.
- 980 Hakim, G. and Masanam, S.: Dynamical tests of a deep-learning weather prediction model, *Artificial Intelligence for the Earth Systems*, 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, e2023GL104 928, <https://doi.org/10.1029/2023GL104928>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2023GL104928>, 2023.
- 985 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, 220 061, 2023.
- Hazeleger, W., Aerts, J., Bauer, P., Bierkens, M., Camps-Valls, G., Dekker, M., Doblas-Reyes, F., and Eyring, V. e. a.: Digital twins of the Earth with and for humans, *Communications Earth & Environment*, 5, 463, 2024.
- 990 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, 2023.
- 995 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](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 PM<sub>2.5</sub> Exposure Assessment in China from 2013 to 2019, *Environmental Science & Technology*, 55, 2152–2162, <https://doi.org/10.1021/acs.est.0c05815>, publisher: American Chemical Society, 2021.
- 1000 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, e2023JD039 202, 2024.
- 1005 Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Fleming, 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, *Atmospheric Chemistry and Physics*, 19, 3515–3556, <https://doi.org/10.5194/acp-19-3515-2019>, publisher: Copernicus GmbH, 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, *Atmospheric Chemistry and Physics*, 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>, number: 21 Publisher: Multidisciplinary Digital Publishing Institute, 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, *Atmospheric Measurement Techniques*, 11, 3457–3477, <https://doi.org/10.5194/amt-11-3457-2018>, publisher: Copernicus GmbH, 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, *Atmospheric Measurement Techniques*, 16, 2627–2640, <https://doi.org/10.5194/amt-16-2627-2023>, publisher: Copernicus GmbH, 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<sub>2</sub> and O<sub>3</sub> concentrations using TROPOMI data and machine learning over East Asia, *Environmental Pollution*, 288, 117711, <https://doi.org/10.1016/j.envpol.2021.117711>, 2021.
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L.: Physics-informed machine learning, *Nature Reviews Physics*, 3, 422–440, <https://doi.org/10.1038/s42254-021-00314-5>, tex.copyright: 2021 Springer Nature Limited, 2021.
- Kazemi Garajeh, M., Laneve, G., Rezaei, H., Sadeghnejad, M., Mohamadzadeh, N., and Salmani, B.: Monitoring trends of CO, NO<sub>2</sub>, SO<sub>2</sub>, and O<sub>3</sub> 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, *Geoscientific Model Development*, 12, 1209–1225, <https://doi.org/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>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021MS002926>, 2022.



- 1045 Kim, M., Brunner, D., and Kuhlmann, G.: Importance of satellite observations for high-resolution mapping of near-surface NO<sub>2</sub> by machine learning, *Remote Sensing of Environment*, 264, 112 573, <https://doi.org/10.1016/j.rse.2021.112573>, 2021.
- Kleinert, F., Leufen, L., and Schultz, M.: IntelliO<sub>3</sub>-ts v1.0: A neural network approach to predict near-surface ozone concentrations in Germany, *Geoscientific Model Development*, 14, 1–25, 2021.
- Kolehmainen, M., Martikainen, H., and Ruuskanen, J.: Neural networks and periodic components used in air quality forecasting, *Atmospheric Environment*, 35, 815–825, 2001.
- 1050 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.
- 1055 Kuo, C.-P. and Fu, J. S.: Ozone response modeling to NO<sub>x</sub> 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 NO<sub>2</sub> 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.
- 1060 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>, section: Journal of Atmospheric and Oceanic Technology, 2021.
- 1065 Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirmsberger, 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>, <https://www.science.org/doi/pdf/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.: 1070 Tropospheric ozone assessment report: Global ozone metrics for climate change, human health, and crop/ecosystem research, *Elem Sci Anth*, 6, 27, 2018.
- Lelieveld, J. and Dentener, F. J.: What controls tropospheric ozone?, *Journal of Geophysical Research: Atmospheres*, 105, 3531–3551, <https://doi.org/10.1029/1999JD901011>, [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/1999JD901011](https://onlinelibrary.wiley.com/doi/pdf/10.1029/1999JD901011), 2000.
- Lessig, C., Luise, I., Gong, B., Langguth, M., Stadler, S., and Schultz, M.: AtmoRep: A stochastic model of atmosphere dynamics using 1075 large scale representation learning, *arXiv preprint arXiv:2308.13280*, 2023.
- Leufen, L. H., Kleinert, F., and Schultz, M. G.: O<sub>3</sub>ResNet: 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 1080 satellite-retrieved atmospheric composition data: OMI SO<sub>2</sub> as an example, *Atmospheric Measurement Techniques*, 15, 5497–5514, <https://doi.org/10.5194/amt-15-5497-2022>, publisher: Copernicus GmbH, 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.
- 1085 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>, number: 5 Publisher: Multidisciplinary Digital Publishing Institute, 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.
- 1090 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, *Atmospheric Chemistry and Physics*, 22, 12543–12557, <https://doi.org/10.5194/acp-22-12543-2022>, 2022a.
- 1095 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*, pp. 11976–11986, [https://openaccess.thecvf.com/content/CVPR2022/html/Liu\\_A\\_ConvNet\\_for\\_the\\_2020s\\_CVPR\\_2022\\_paper.html](https://openaccess.thecvf.com/content/CVPR2022/html/Liu_A_ConvNet_for_the_2020s_CVPR_2022_paper.html), 2022b.
- Liu, Z.-S., Clusius, P., and Boy, M.: Neural Network Emulator for Atmospheric Chemical ODE, <https://doi.org/10.48550/arXiv.2408.01829>, 2024.
- 1100 Loyola, D. G., Koukouli, M. E., Valks, P., Balis, D. S., Hao, N., Van Roozendaal, M., Spurr, R. J. D., Zimmer, W., Kiemle, S., Lerot, C., and Lambert, J.-C.: The GOME-2 total column ozone product: Retrieval algorithm and ground-based validation, *Journal of Geophysical Research: Atmospheres*, 116, <https://doi.org/10.1029/2010JD014675>, <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2010JD014675>, 2011.
- 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, *Atmospheric Chemistry and Physics*, 19, 8339–8361, <https://doi.org/10.5194/acp-19-8339-2019>, publisher: Copernicus GmbH, 2019.
- 1105 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](https://doi.org/10.1016/S2542-5196(22)00260-1), publisher: Elsevier, 2022.
- 1110 Malley, C. S., Henze, D. K., Kuylensstierna, 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  $\geq 30$  Years 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, <https://doi.org/10.48550/arXiv.2406.16947>, arXiv:2406.16947, 2024.
- 1115 Marécal, V., Peuch, V., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., and Chéroux, F.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, *Geoscientific Model Development*, 8, 2777–2813, 2015.



- 1120 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>, publisher: Taylor & Francis \_eprint: <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, 1125 <https://doi.org/10.1111/gcb.14157>, \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.14157>, 2018.
- Miyazaki, K., Bowman, K., Sekiya, T., Eskes, H., Boersma, F., Worden, H., Livesey, N., Payne, V. H., Sudo, K., Kanaya, Y., Takigawa, M., and Ogochi, K.: Updated tropospheric chemistry reanalysis and emission estimates, TCR-2, for 2005–2018, *Earth System Science Data*, 12, 2223–2259, <https://doi.org/10.5194/essd-12-2223-2020>, publisher: Copernicus GmbH, 2020a.
- Miyazaki, K., Bowman, K. W., Yumimoto, K., Walker, T., and Sudo, K.: Evaluation of a multi-model, multi-constituent assimilation frame- 1130 work for tropospheric chemical reanalysis, *Atmospheric Chemistry and Physics*, 20, 931–967, <https://doi.org/10.5194/acp-20-931-2020>, publisher: Copernicus GmbH, 2020b.
- ML4O3: ML4O3/Applications-of-Machine-Learning-and-Artificial-Intelligence-in-Tropospheric-Ozone-Research, <https://github.com/ML4O3/Applications-of-Machine-Learning-and-Artificial-Intelligence-in-Tropospheric-Ozone-Research>, original-date: 2024-11-19T12:37:00Z, 2024.
- 1135 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 Schneidmesser, 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, *Atmospheric Chemistry and Physics*, 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., 1140 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), *Geoscientific Model Development*, 10, 639–671, <https://doi.org/10.5194/gmd-10-639-2017>, 2017.
- 1145 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>, tex.copyright: © 2023. The Authors., 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, 1150 and Implementation, <https://doi.org/10.48550/arXiv.2309.10808>, 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>, \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2002JD002784>, 2003.
- 1155 Naeger, A. R. e. a.: Revolutionary air-pollution applications from future tropospheric emissions: monitoring of pollution (TEMPO) observations, *Bulletin of the American Meteorological Society*, 102, E1735–E1741, 2021.



- Nanda, N., Chan, L., Lieberum, T., Smith, J., and Steinhardt, J.: Progress measures for grokking via mechanistic interpretability, arXiv preprint arXiv:2301.05217, 2023.
- National Research Council (U.S.): A National Strategy for Advancing Climate Modeling, National Academies Press, <https://doi.org/10.17226/13430>, 2012.
- 1160 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>, publisher: Nature Publishing Group, 2014.
- 1165 Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., and Grover, A.: ClimaX: A foundation model for weather and climate, arXiv preprint 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, 104 016, <https://doi.org/10.1088/1748-9326/aae2be>, 2018.
- 1170 Nowack, P., Runge, J., Eyring, V., and Haigh, J.: Causal networks for climate model evaluation and constrained projections, *Nature Communications*, 11, 1415, 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, *Atmospheric Measurement Techniques*, 17, 5147–5159, <https://doi.org/10.5194/amt-17-5147-2024>, publisher: Copernicus GmbH, 2024.
- 1175 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, 22 513, <https://doi.org/10.1038/s41598-021-01824-z>, tex.copyright: 2021 The Author(s), 2021.
- 1180 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), pp. 1–4, <https://doi.org/10.1109/CASAP54985.2021.9703381>, 2021.
- 1185 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, <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.
- 1190 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, <https://doi.org/10.48550/arXiv.2312.15796>, arXiv:2312.15796, 2024.



- Prospero, J. M., Ginoux, P., Torres, O., Nicholson, S. E., and Gill, T. E.: Environmental characterization of global sources of atmospheric soil dust identified with the nimbus 7 total ozone mapping spectrometer (toms) absorbing aerosol product, *Reviews of Geophysics*, 40, 1195  
<https://doi.org/10.1029/2000RG000095>, 2002.
- 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>, publisher: Proceedings of the National Academy of Sciences, 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, 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, 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>, number: 7743 Publisher: Nature Publishing Group, 2019.
- 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 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](https://doi.org/10.1016/S1352-2310(97)83212-8), 1997.
- Savage, N., Agnew, P., Davis, L., Ordóñez, C., Thorpe, R., Johnson, C., O’Connor, F., and Dalvi, M.: Air quality modelling using the Met Office Unified Model (AQUUM OS24-26): model description and initial evaluation, *Geoscientific Model Development*, 6, 353–372, 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>, number: 1 Publisher: Nature Publishing Group, 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>, tex.copyright: © 2022 The Authors. *Journal of Advances in Modeling Earth Systems* published by Wiley Periodicals LLC on behalf of American Geophysical Union., 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.



- 1230 Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., and Stadler, 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>, publisher: Royal Society, 2021.
- Sengupta, P., Zhang, Y., Maharjan, S., and Eliassen, F.: Balancing Explainability-Accuracy of Complex Models, <https://doi.org/10.48550/arXiv.2305.14098>, 2023.
- 1235 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<sub>2</sub> measurements, *Atmospheric Chemistry and Physics*, 23, 1227–1257, <https://doi.org/10.5194/acp-23-1227-2023>, publisher: Copernicus GmbH, 2023.
- 1240 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, *Geoscientific Model Development*, 13, 2475–2486, <https://doi.org/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, *Geoscientific Model Development*, 15, 1677–1687, <https://doi.org/10.5194/gmd-15-1677-2022>, 2022.
- 1245 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.
- 1250 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>, tex.copyright: ©2018. American Geophysical Union. All Rights Reserved., 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., Oliví, 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.
- 1260 Stephanie Russo Carroll et al.: Operationalizing the CARE and FAIR principles for indigenous data futures, *Scientific Data*, 8, 108, 2021.
- Sturm, P. O. and Wexler, A. S.: A mass- and energy-conserving framework for using machine learning to speed computations: a photochemistry example, *Geoscientific Model Development*, 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), *Geoscientific Model Development*, 15, 3417–3431, <https://doi.org/10.5194/gmd-15-3417-2022>, 2022.
- 1265 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>, tex.copyright: © 2023 The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union., 2023.
- 1270 Tesch, T., Kollet, S., and Garcke, J.: Causal deep learning models for studying the Earth system, *Geoscientific Model Development*, 16, 2149–2166, 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.
- 1275 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 NO<sub>2</sub> emission strengths over Riyadh and Madrid from space from a combination of wind-assigned anomalies and a machine learning technique, *Atmospheric Measurement Techniques*, 16, 2237–2262, <https://doi.org/10.5194/amt-16-2237-2023>, publisher: Copernicus GmbH, 2023.
- University, I. E. S. I. N. C. C.: Gridded population of the world, version 4 (GPWv4): Population count, revision 11, <https://doi.org/10.7927/H4JW8BX>, 2018.
- 1280 Veefkind, J., de Haan, J., Brinksma, E., Kroon, M., and Levelt, P.: Total ozone from the ozone monitoring instrument (OMI) using the DOAS technique, *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1239–1244, <https://doi.org/10.1109/TGRS.2006.871204>, conference Name: IEEE Transactions on Geoscience and Remote Sensing, 2006.
- 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.
- 1285 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.
- 1290 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>, \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021JD036181>, 2022.
- 1295 Wilkinson, M., M, D., and al., A. I. e.: The FAIR Guiding Principles for scientific data management and stewardship, *Nature Scientific Data*, 160018, 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, *Geoscientific Model Development*, 16, 1801–1821, <https://doi.org/10.5194/gmd-16-1801-2023>, 2023.
- 1300 Wu, B., Wu, C., Ye, Y., Pei, C., Deng, T., Li, Y., Lu, X., Wang, L., Hu, B., Li, M., and Wu, D.: Long-term hourly air quality data bridging of neighboring sites using automated machine learning: A case study in the Greater Bay area of China, *Atmospheric Environment*, p. 120347, 2024.
- 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, <https://www.authorea.com/users/816476/articles/>



- 1305 1217166-advancing-photochemistry-simulation-in-wrf-chem-v4-0-artificial-intelligence-photochemistry-aipc-scheme-with-multi-head-self-attention  
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<sub>2.5</sub> prediction models, *Atmospheric Environment*, 244, 117921, <https://doi.org/10.1016/j.atmosenv.2020.117921>, 2021.
- 1310 Xie, F., Ren, T., Zhao, C., Wen, Y., Gu, Y., Zhou, M., Wang, P., Shiomi, K., and Morino, I.: Fast retrieval of XCO<sub>2</sub> over east Asia based on Orbiting Carbon Observatory-2 (OCO-2) spectral measurements, *Atmospheric Measurement Techniques*, 17, 3949–3967, <https://doi.org/10.5194/amt-17-3949-2024>, publisher: Copernicus GmbH, 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.
- 1315 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>, publisher: American Chemical Society, 2022.
- Yafouz, A., AlDahoul, N., Birima, A. H., Ahmed, A. N., Sherif, M., Sefelnasr, A., Allawi, M. F., and Elshafie, A.: Comprehensive comparison  
1320 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, <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,  
1325 <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.
- 1335 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>, \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/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, *Geoscientific Model Development*, 16, 199–209, <https://doi.org/10.5194/gmd-16-199-2023>, publisher: Copernicus GmbH, 2023.
- 1340



- 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.
- 1345 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, *Atmospheric Chemistry and Physics*, 19, 3257–3269, <https://doi.org/10.5194/acp-19-3257-2019>, publisher: Copernicus GmbH, 2019.
- 1350 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.