Deep Learning-Based Ensemble Forecasts and Predictability Assessments for Surface Ozone Pollution

Aoxing Zhang[1,2] [ID], Tzung-May Fu[3] [ID], Xu Feng[4], Jianfeng Guo[5], Chanfang Liu[5], Jiongkai Chen[1,2], Jiajia Mo[1,2], Xiao Zhang[6], Xiaolin Wang[7] [ID], Wenlu Wu[1,2], Yue Hou[1,2], Honglong Yang[8], and Chao Lu[8]

[1]Shenzhen Key Laboratory of Precision Measurement and Early Warning Technology for Urban Environmental Health Risks, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China, [2]Guangdong Provincial Observation and Research Station for Coastal Atmosphere and Climate of the Greater Bay Area, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China, [3]National Center for Applied Mathematics, Shenzhen (NCAMS), Shenzhen, China, [4]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA, [5]Shenzhen Ecology and Environment Monitoring Centre of Guangdong Province, Shenzhen, China, [6]Department of Computer Science and Technology, Tsinghua University, Beijing, China, [7]Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China, [8]Shenzhen National Climate Observatory, Shenzhen, China

Abstract The impacts of weather forecast uncertainties have not been quantified in current air quality forecasting systems. To address this, we developed an efficient 2-D convolutional neural network-surface ozone ensemble forecast (2DCNN-SOEF) system using 2-D convolutional neural network and weather ensemble forecasts, and we applied the system to 216-hr ozone forecasts in Shenzhen, China. The 2DCNN-SOEF demonstrated comparable performance to current operating forecast systems and met the air quality level forecast accuracies required by the Chinese authorities up to 144-hr lead time. Uncertainties in weather forecasts contributed 38%–54% of the ozone forecast errors at 24-hr lead time and beyond. The 2DCNN-SOEF enabled an "ozone exceedance probability" metric, which better represented the risks of air pollution given the range of possible weather outcomes. Our ensemble forecast framework can be extended to operationally forecast other meteorology-dependent environmental risks globally, making it a valuable tool for environmental management.

Plain Language Summary Weather forecasts are intrinsically uncertain, but the impacts of that uncertainty on air quality forecasts are not explicitly quantified in current air quality forecast systems. We proposed here a surface ozone ensemble forecast system, analogous to modern weather ensemble forecast systems, to represent the probability distribution of forecasted surface ozone concentrations given 30–50 possible future weather outcomes. The computation costs of this surface ozone ensemble forecast system were greatly reduced using deep learning techniques that emphasized the spatial patterns of weather. We showed that the surface ozone ensemble forecast system's accuracy met the Chinese operational requirements. However, half of the ozone forecast error was due to weather forecast uncertainties, which cannot be completely eliminated even with perfect pollutant emission estimates and chemistry models. This weather-induced innate uncertainty in air quality forecasts should be considered for effective air quality management.

1. Introduction

Surface ozone is the dominant warm-season air pollutant in many global cities (Ministry of Ecology and Environment of the People's Republic of China (MEEC), 2021; U.S. Environmental Protection Agency (USEPA), 2021) and was responsible for 365,000 premature mortality worldwide in 2019 (Murray et al., 2020). In response to the health risks of ozone pollution, many cities have promulgated emergency response measures by abating precursor emissions or by issuing public health advisories when surface ozone concentrations are forecasted to exceed the local air quality standards (Ministry of Environmental Protection of the People's Republic of China, 2015; USEPA, 2015).

The effective implementation of these emergency response measures hinges on the accurate forecasts of ozone exceedances several days in advance. For instance, the MEEC requires that air quality forecasts be issued daily at the city level for the future 5 days, and that the daily air quality level forecasts (defined in Text S1 in Supporting Information S1) for the future 1–3 days be correct ≥60% of the days in a year (MEEC, 2020). In addition,

the MEEC requires that the Air Quality Index (AQI) forecasts achieve accuracies of ± 10 for non-exceedance days and ± 15 for light pollution days (corresponding to ± 12 and ± 16.5 μg m^{-3} for maximum daily 8-hr average (MDA8) ozone concentration forecasts, respectively; or ± 8 and ± 30 μg m^{-3} of hourly ozone concentration forecasts, respectively) (Text S1 in Supporting Information S1). However, a previous study showed that the median root-mean-square error (RMSE) of hourly ozone concentration forecasts in 34 Chinese cities from seven operational 72-hr forecast systems was approximately 40 μg m^{-3} in summer with regional differences (Petersen et al., 2019), not always compliant with the MEEC's required accuracy. What factors limit the accuracy of surface ozone forecasts, and whether the mandated forecast accuracies can be achieved using current forecasting tools, have not been systematically evaluated.

Surface ozone pollution is often correlated with meteorological conditions that favor the photochemical production of ozone and the accumulation of its precursors, such as high temperature, strong solar radiation, surface wind convergence, and boundary-layer stagnation (Fu & Tian, 2019; Jacob & Winner, 2009). These local conditions are in turn caused by various synoptic to mesoscale weather. In the Pearl River Delta (PRD) area of Southern China, for example, ozone pollution events are often associated with the regional subsidence or surface wind convergence induced by the west Pacific subtropical high (Jiang et al., 2021), the outflow of an approaching typhoon (e.g., Li et al., 2022; Ouyang et al., 2022), or the land-sea breeze (Ding et al., 2004). As such, the uncertainty in synoptic (e.g., typhoon track) to mesoscale (e.g., land-sea breeze) weather forecasts is a substantial source of uncertainty to surface ozone forecasts (Lam et al., 2018; N. Wang et al., 2022). However, current operational air quality forecasts are based on a small number of regional air quality simulations and cannot represent the range of possible future weather outcomes. Moreover, weather forecast uncertainties increase with lead time (the length of time between the issuance of a forecast and the occurrence of the predicted event) (Slingo & Palmer, 2011) and may pose a "predictability" limit to air quality; that is, air quality forecasts beyond a certain lead time may not be sufficiently accurate to satisfy requirements or managerial needs.

The uncertainties of weather forecasts stem fundamentally from the non-linearity of atmospheric dynamics and its intrinsic sensitivity to initial conditions (Ehrendorfer, 1997), that is, a small perturbation in the initial conditions would lead to large and growing deviation in the system's behavior. In modern weather forecasting systems, forecast uncertainty is quantified by an ensemble of 30–50 forecast members; each member is a model realization with slightly perturbed initial conditions or physical parameters, such that the ensemble members span the range of possible weather outcomes. It stands to reason that the meteorological uncertainty of surface ozone forecasts should also be quantified with an ensemble approach. However, 3-D regional air quality models are computationally expensive, such that conducting many simulations with perturbed weather forecasts is not realistic for daily operations.

We propose here the use of machine learning/deep learning (ML/DL) methods to efficiently conduct surface ozone ensemble forecasts and quantify the meteorological uncertainty. Previous studies have demonstrated the success of ML/DL methods in air quality forecasts (e.g., Athira et al., 2018; Sayeed et al., 2020, 2021; Sun et al., 2021). However, previous ML/DL models mostly trained with locally observed meteorological variables and pollutant concentrations, thus removing the spatial information associated with synoptic to mesoscale weather and may be affected by pollutant observations during emergency emission abatement events. Moreover, precursor emissions have been changing rapidly in China and other developing countries, such that the ML/DL models trained with historic measurements may not reflect current precursor forcing.

In this study, we combined the 2-D convolutional neural network (2DCNN) method, which emphasized spatial patterns (Huang et al., 2021; Xing et al., 2020), and weather ensemble forecasts to construct a surface ozone ensemble forecast system (2DCNN-SOEF), with the goal of quantifying the meteorological uncertainties of ozone forecasts. We generated a large training data set by perturbing a regional air quality model with a wide range of synoptic-to-mesoscale meteorological variations. As a proof of concept, we constructed a daily 216-hr 2DCNN-SOEF system for Shenzhen City in the PRD area of China. We evaluated the skills and forecast uncertainties of the 2DCNN-SOEF system against observations during two ozone pollution seasons.

2. Materials and Methods

2.1. Surface Ozone Simulations Using the WRF-GC Model

We used the WRF-GC v2.0 regional air quality model (Feng et al., 2021; Lin et al., 2020) to simulate surface ozone over China during Shenzhen's main ozone pollution season (25 June to 6 October with one spin-up day) of
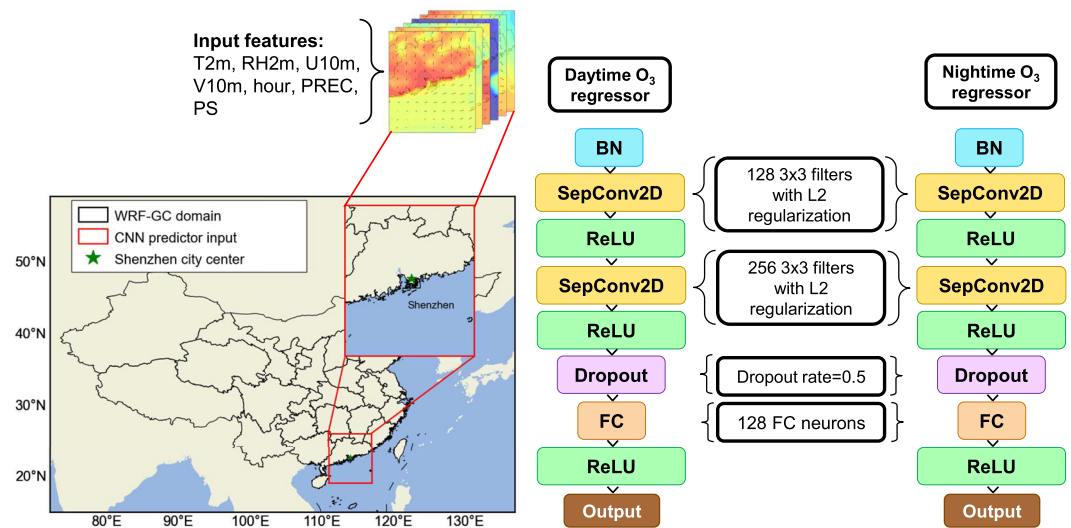
Figure 1. Schematic of the 2-D convolutional neural network (2DCNNs) trained with Weather Research and Forecasting (WRF)-GC simulated meteorological fields and surface ozone concentrations. The map shows the WRF-GC simulation domain (black box), the spatial extent of the meteorological fields used for driving the 2DCNNs (red box), and the model grid of Shenzhen city center where surface ozone concentrations were predicted (star). The right panel shows the structure of the 2DCNNs. The predictor variables included 2-m air temperature (T2m), 2-m relative humidity (RH2m), 10-m zonal and meridional winds (U10m and V10m), 6-hr accumulated precipitation (PREC), surface pressure (PS), and the hour-of-the-day (hour).

the years 2018, 2019, and 2021; the simulated results around Shenzhen (red box in Figure 1) were used as training and evaluation data sets for the 2DCNNs. WRF-GC is an online coupling of the Weather Research and Forecasting (WRF) meteorological model (v3.9.1.1) (Skamarock et al., 2008, 2019) and the GEOS-Chem chemical transport model (v12.8.2) (Bey et al., 2001). Figure 1 shows our simulation domain (16.4°–54.7°N, 72.8°–137.2°E), with 27 km horizontal resolution and 49 vertical levels. Further details of the WRF-GC model and our emissions and physical settings are described in Text S2 in Supporting Information S1.

Text S2 and Table S1 in Supporting Information S1 describes our WRF-GC simulations. Briefly, the "WG-FNL-2021" simulation was driven and nudged with the meteorological reanalysis data from the National Centers for Environmental Prediction Final Operational Global Analysis (NCEP GDAS/FNL, 0.25° spatial resolution, https://rda.ucar.edu/datasets/ds083.3/ last accessed: 6 May 2022; NCEP, 2000) for summer 2021, such that the simulated meteorological fields were as close to reality as possible. In order to generate a large training data set for the 2DCNNs, we conducted the "WG-FNL-2018" and "WG-hindcast-2021" simulations to represent the ozone-meteorology relationship of the years 2018 and 2021. These simulations were designed with the following goals: (a) representing a wide range of realistic synoptic weather affecting surface ozone, and (b) representing the mesoscale meteorological variations affecting surface ozone under those typical synoptic weather scenarios. These simulations produced 16,464 hourly WRF-GC outputs, which we randomly divided into a training set (85%, 13,994 hr) and a test set (15%, 2470 hr). We further conducted a "WG-FNL-2019" simulation for an independent year (2019) and used the results (2,496 hr) to validate the 2DCNNs (Weng et al., 2022). Finally, we drove and nudged the WRF-GC model with 10 members from the ensemble data assimilation (EDA) (Isaksen et al., 2010) of the European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ECMWF ERA5 https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5, last accessed: 6 May 2022; Hersbach et al., 2020) for 2021. Each of the EDA members differed only in meso- to small-scale meteorology; the WRF-GC simulations driven by those members ("WG-EDA-2021") were used to evaluate whether the 2DCNNs could discern surface ozone variations due to small differences in meteorology.

2.2. Construction of the 2DCNNs for Surface Ozone Prediction in Shenzhen

We designed 2DCNNs to predict hourly surface ozone concentrations at Shenzhen city center (22.36°–22.65°N, 113.93°–114.24°E). Figure 1 shows the structure of our 2DCNN models; Text S3 in Supporting Information S1

describes the layers and parameters. The 2DCNNs were trained with WRF-GC-simulated hourly meteorological fields over Guangdong Province and its coastal waters (the 648 km × 513 km red box in Figure 1) and the WRF-GC-simulated surface ozone concentrations at that hour at the model grid of Shenzhen city center, such that the 2DCNNs may capture synoptic to mesoscale meteorological features and the influences of both continental and marine air masses. We trained the 2DCNNs to predict ozone concentrations of any given hour using meteorological conditions of that hour. In this way, the 2DCNNs' ozone predictions were implicitly dependent on meteorological histories before the predicted hour, because weather conditions evolved continuously. We trained separate 2DCNNs for daytime (8:00 to 18:00 local time) and nighttime (19:00 to 7:00 local time) to represent the impacts of daytime photochemistry and nocturnal boundary layer dynamics on surface ozone concentrations. Guided by our experience in operational weather and air quality forecasts in Shenzhen and by trial-and-error, we selected the following hourly meteorological predictors: 2-m air temperature (T2m) and relative humidity (RH2m), 10 m zonal (U10m) and meridional (V10m) wind, surface pressure (PS), 6-hr accumulated precipitation (PREC), and an additional local time-of-day variable (hour, trigonometrically transformed following Sun and Archibald (2021)). Low planetary boundary layer heights (PBLHs) often occur under subsidence and stagnant conditions (Dong et al., 2020) and would be a strong predictor for surface ozone pollution. However, meteorological models often have different vertical resolutions and define PBLH differently, such that the inclusion of PBLH would reduce the generalizability of the 2DCNNs to various meteorological data sets. Because low PBLHs often correlate with high T2m, low RH2m, high PS, and low wind speed, we relied on these predictors to implicitly represent the impacts of low PBLH in the 2DCNNs.

2.3. Hindcasts and Forecasts of Surface Ozone in Shenzhen Using the 2DCNNs

The performance of our trained 2DCNNs was evaluated in a series of ozone "hindcasts" (driven by meteorological reanalysis data) and "forecasts" (driven by actual weather forecasts) during the ozone seasons of 2019 and 2021 (Text S4 in Supporting Information S1). For driving the 2DCNNs, the input meteorological fields were first interpolated to the 27 km-resolution grids of our WRF-GC model using a cubic spline. We drove the 2DCNN hindcasts with meteorological reanalyses from the NCEP GDAS/FNL and the ERA5 for the years 2019 and 2021, respectively. The use of meteorological data from different years assessed the 2DCNNs' performance under different weather scenarios, while the use of two reanalysis data sets assessed the 2DCNNs' robustness against different meteorological data sources. Finally, we drove the 2DCNNs with the ECMWF 216-hr 50-member ensemble weather forecast (0.125° resolution, 3-hr temporal resolution, https://www.ecmwf.int/en/forecasts/datasets/set-iii, last accessed: 26 May 2022; Molteni et al., 1996; Palmer et al., 1997) during Shenzhen's ozone season of 2019 and 2021 ("2DCNN-SOEF-2019" and "2DCNN-SOEF-2021" in Table S1 in Supporting Information S1). These 2DCNN ensemble surface ozone forecasts (2DCNN-SOEFs) were used to quantify the uncertainty of surface ozone forecasts as a result of actual weather forecast uncertainties.

2.4. Hourly Surface Ozone Measurements

We evaluated the WRF-GC ozone simulations and the 2DCNN-SOEF predictions against quality-controlled (X. Wang et al., 2021) hourly ozone measurements from the China National Environmental Monitoring Centre (http://www.cnemc.cn, last accessed: 5 May 2022). We averaged hourly measurements at the seven sites within the metropolitan area (Table S4 in Supporting Information S1) to represent the surface ozone levels in Shenzhen.

3. Results

3.1. Evaluation of the Ozone-Meteorology Relationship Simulated by the WRF-GC Model and Its Manifestation in the 2DCNNs

We first evaluated whether WRF-GC correctly represented the surface ozone-meteorology relationship at our location of interest. The WG-FNL-2021 simulation nudged with observed meteorology reproduced the observed surface ozone concentrations over Southern China during late June to early October 2021 (Figure S2 in Supporting Information S1; spatial correlation $r = 0.81$, normalized mean bias NMB $= 0.15 \pm 0.12$). The simulated time series of MDA8 surface ozone concentrations in Shenzhen were highly consistent with the observations (Figure S3 in Supporting Information S1; temporal correlation $r = 0.82$, NMB $= 0.10 \pm 0.38$) during that period, indicating that the WRF-GC model, using with our emissions and physical settings, represented the relationship between synoptic meteorology and surface ozone concentrations in Shenzhen.

Based on WRF-GC's good performance in simulating surface ozone in Shenzhen, we trained 2DCNNs using WRF-GC simulations to predict surface ozone concentrations at Shenzhen city center. Figure S4 in Supporting Information S1 evaluates the fitting of our 2DCNNs, that is, the consistency between hourly ozone concentrations calculated by the 2DCNNs driven by WRF-GC meteorological simulations and the hourly ozone concentrations directly simulated by WRF-GC. The 2DCNNs showed good agreement with WRF-GC simulations ($r = 0.98$, NMB = 0.007 for the training set; $r = 0.98$, NMB = 0.002 for the test set; $r = 0.91$, NMB = -0.131 for the validation using WRF-GC results from the independent year, 2019). In particular, in the independent validation set, 60% of the hourly surface ozone exceedances (>200 µg m^{-3}, Text S1 in Supporting Information S1) simulated by WRF-GC were successfully predicted by the 2DCNNs ("2DCNN-WG-2019", Text S4 and Table S1 in Supporting Information S1). Altogether, 98% of the hourly surface ozone exceedance and non-exceedance events simulated by WRF-GC were successfully predicted by the 2DCNNs. We concluded that the 2DCNNs captured the ozone-meteorology relationship represented by our WRF-GC simulations for both low- and high-ozone cases. Of the six meteorological predictors used in the 2DCNNs, the 10-m meridional wind (V10m) was the most important predictor of high surface ozone concentrations in Shenzhen (Text S5 and Figure S5 in Supporting Information S1); this sensitivity reflected the transport of ozone and its precursors from the inland cities to Shenzhen by northerly winds and the transport of cleaner marine air from the South China Sea to Shenzhen by southerly winds, consistent with our operational experience and previous studies (e.g., Ding et al., 2004).

We experimented with other deep neural networks or multiple linear regression models driven only by local meteorological variables. We found that the 2DCNN framework better captured the ozone-meteorology relationship, as its spatial convolution layers accentuated the spatial connections between ozone and weather patterns (Figure S6 in Supporting Information S1). We also found that mesoscale meteorological differences (e.g., mesoscale convergence of surface winds shown in Figure S7 in Supporting Information S1) may lead to ± 50 µg m^{-3} variation in the hourly surface ozone concentrations simulated by WRF-GC, consistent with previous studies (Li et al., 2022; Zhang et al., 2007); that sensitivity of surface ozone to mesoscale meteorology was reproduced by the 2DCNNs (Text S5 in Supporting Information S1).

We next applied the 2DCNNs to "hindcast" surface ozone concentrations in Shenzhen. Figure 2 shows the surface ozone concentration hindcasts from the 2DCNNs driven with different meteorological reanalysis data sets and compares them against surface observations. We focused our analyses on the results at 14:00 local time (6:00 UTC), when surface ozone concentrations were often highest in a day. The 2DCNN hindcasts driven by different meteorological reanalysis data sets consistently showed good agreement with observations in 2019 and 2021 ($r = 0.73$ to 0.91, with no systematic biases), demonstrating that the 2DCNNs were flexible with regard to the sources of the meteorological data set and that the 2DCNN's representation of ozone-meteorology relationship was robust. For the year 2019 (not used for training the 2DCNNs), the hindcasts driven by different reanalysis data captured 60% (2DCNN-ERA5-2019) and 91% (2DCNN-FNL-2019) of the observed hourly ozone exceedance events at 14:00 local time, as well as 60% (2DCNN-ERA5-2019) to 76% (2DCNN-FNL-2019) of the observed MDA8 ozone exceedance events.

3.2. Performance of the 2DCNN-Surface Ozone Ensemble Forecasts (2DCNN-SOEFs)

We drove the 2DCNNs with the ECMWF 50-member meteorological ensemble forecasts at 24 and 72-hr lead time, to produce an ensemble of surface ozone concentration forecasts given the range of possible future weather outcomes. Figure 2 shows the results from these 2DCNN-SOEFs during late June to early October of 2019 and 2021 (2DCNN-SOEF-2019 and 2DCNN-SOEF-2021 in Text S4). At 24-hr lead time, the ensemble means of the surface ozone concentration forecasts at 14:00 local time were in good agreement with the observations for both 2019 ($r = 0.84$, NMB = -0.11 ± 0.33) and 2021 ($r = 0.83$, NMB = 0.03 ± 0.36). The ensemble mean daily air quality level forecasts were correct on 75% and 90% of the forecasted days in 2019 and 2021, respectively. The skills of the 2DCNN-SOEF deteriorated slightly at 72-hr lead time (ensemble mean vs. observation $r = 0.81$, NMB = -0.14 ± 0.35 in 2019; $r = 0.78$, NMB = 0.02 ± 0.41 in 2021), with the 2DCNN-SOEF correctly forecasting the daily air quality levels on 76% and 87% of the days during the ozone pollution seasons of 2019 and 2021, respectively. Therefore, the 2DCNN-SOEF system met the MEEC's required accuracy for air quality level forecasts (correct air quality level forecast on $>60\%$ of the days) at 1- to 3-day lead time.

The development of the 2DCNN-SOEF also allowed us to express the meteorological uncertainty of ozone pollution forecasts in an "ozone exceedance probability", that is, the percentage of ensemble members predicting
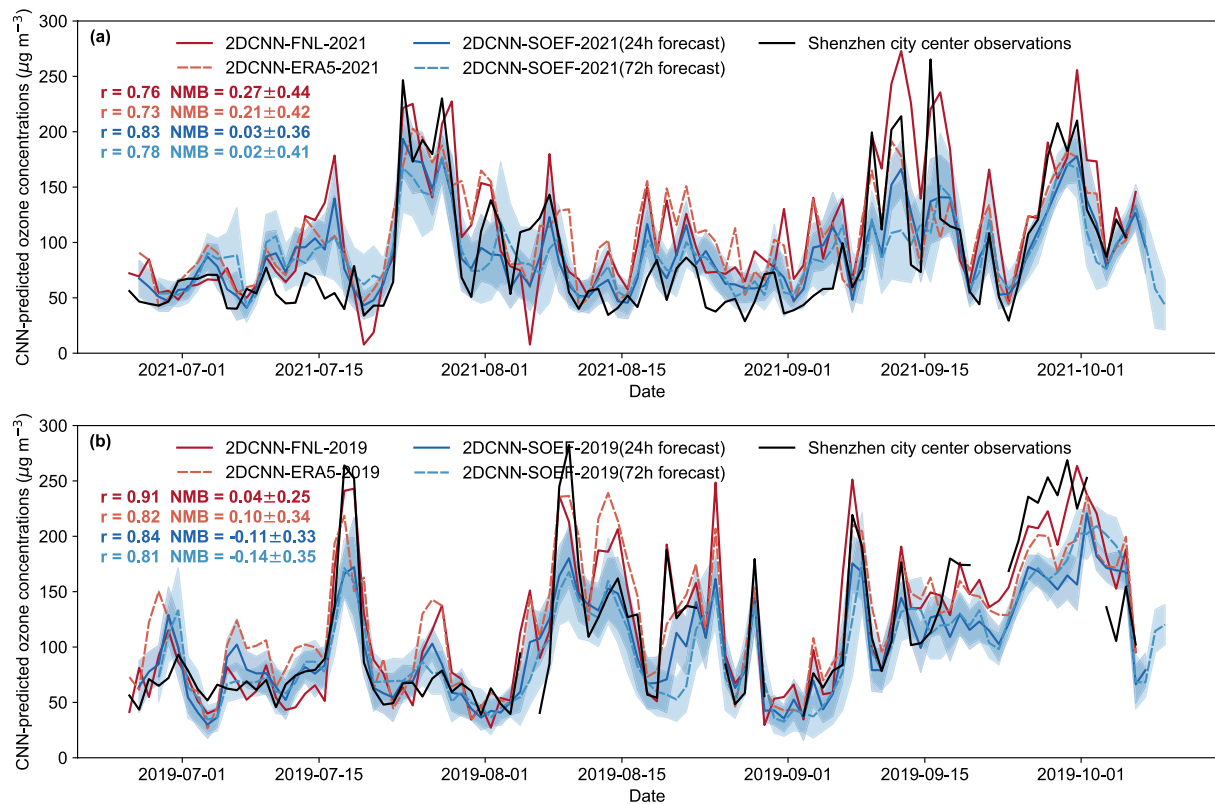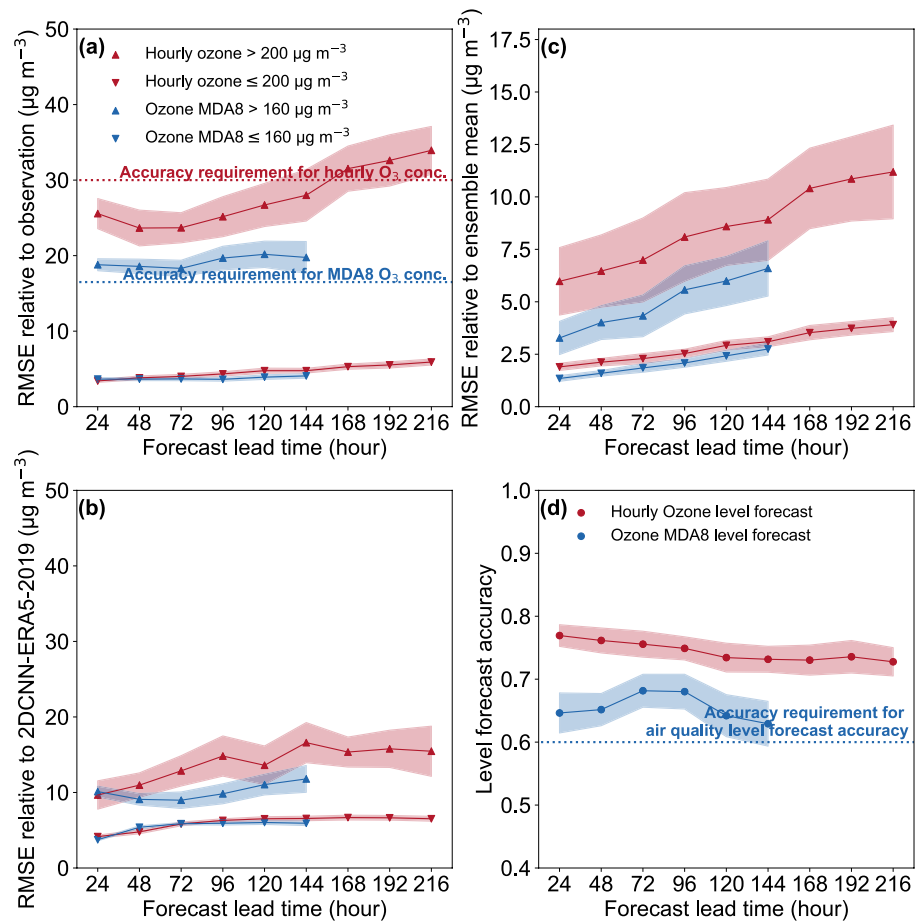
Figure 2. Surface ozone concentrations at Shenzhen at 14:00 local time from the 2-D convolutional neural network (2DCNN) hindcasts and forecasts (color-coded) for late June to early October of (a) 2021 and (b) 2019. The shaded areas indicated the standard deviations of the 2DCNN-surface ozone ensemble forecasts at 24-hr (dark blue) and 72-hr (light blue) lead time, respectively. Also shown are the surface observations at Shenzhen (black line). The correlation coefficients (*r*) and normalized mean biases of the forecasts against the observations are shown inset.

surface ozone exceedance. This statistical metric is analogous to the precipitation probability in modern weather ensemble forecast systems. Figure S8 in Supporting Information S1 shows that, on 19 days during the ozone pollution season of 2019, the ozone exceedance probability forecasts forewarned the possibility of ozone exceedances even when the forecasted ensemble mean ozone concentrations were not in exceedance. In this way, the 2DCNN-SOEF better informed air quality management, because surface ozone concentrations under a wide range of possible future weather outcomes were predicted.

3.3. Predictability of Surface Ozone Concentrations as Limited by Meteorological Forecast Uncertainties

Our 2DCNN-SOEF system allowed us to quantify the changes of ozone forecast accuracy with lead time and the impacts of weather forecast uncertainties on ozone forecast errors. Figure 3 shows the RMSEs of the 2DCNN-SOEFs relative to observations in Shenzhen for the 2019 ozone season. The MDA8 ozone concentrations were only forecasted up to 144-hr lead time because the output frequencies of the ECMWF weather ensemble forecasts were reduced to every 6 hr after that lead time. The RMSEs for forecasted ozone concentrations in exceedance of the national air quality standards (Text S1 in Supporting Information S1) were both larger and grew faster with lead time, compared to the RMSEs for forecasted ozone concentrations not in exceedance. This characteristic was consistent with the high-ozone events being more sensitive to mesoscale meteorology (Figure S5 in Supporting Information S1), which were in turn more difficult to forecast accurately at longer lead times. For forecasted hourly and MDA8 ozone concentrations exceeding the national standards, the RMSEs were 26 ± 2 and 19 ± 1 $\mu g \, m^{-3}$ at 24-hr lead time, respectively (Figure 3a). At 72-hr lead time, the RMSEs for hourly and MDA8 ozone concentrations in exceedance were 24 ± 2 and 18 ± 1 $\mu g \, m^{-3}$, respectively. These ozone forecast errors were comparable to the 20–30 $\mu g \, m^{-3}$ median RMSE for the 72-hr ozone forecasts over PRD cities from seven operational forecasting systems reported by Petersen et al. (2019). The RMSEs of the 2DCNN-SOEF hourly ozone concentrations were smaller than the MEEC's accuracy requirement (± 30 $\mu g \, m^{-3}$) up to 144-hr lead

Figure 3. The 2-D convolutional neural network surface ozone ensemble forecast's (2DCNN-SOEF) performance for hourly (red) and MDA8 (blue) ozone concentrations during the 2019 ozone season as a function of forecast lead time: (a) root-mean-square errors (RMSEs) of the 2DCNN-SOEF relative to observations, (b) RMSEs of the 2DCNN-SOEF relative to the 2DCNN-ERA5-2019 surface ozone hindcast, (c) RMSEs of the 2DCNN-SOEF ensemble members relative to the ensemble means, and (d) the hit rates of air quality level forecast. The shaded areas indicated the standard deviations among the ensemble members. Also shown are the Ministry of Ecology and Environment of the People's Republic of China's (MEEC) forecast accuracy requirements (dashed lines).

time, but the RMSEs of the MDA8 ozone forecasts exceeded the MEEC's accuracy requirement ($\pm 16.5 \ \mu g \ m^{-3}$) even at 24-hr lead time (Figure 3a). For MDA8 and hourly ozone concentrations not exceeding the national standards, the RMSEs of our 2DCNN-SOEF system were below 8 $\mu g \ m^{-3}$ and satisfied the accuracies required by the MEEC throughout the 9-day forecast window (Figure 3a). In terms of the air quality level forecast, the hit rate (the percentage of the number of days with correct air quality level forecast vs. the total number of forecasted days) of the 2DCNN-SOEF was >60% up to 144-hr lead time for MDA8 ozone and up to 216-hr lead time for hourly ozone during the 2019 ozone season (Figure 3d), compliant with MEEC's requirements.

The 2DCNN-SOEF's error and its deteriorating accuracy with lead time were the consequence of several sources of uncertainties, including uncertainties in precursor emissions and photochemistry inherited from the WRF-GC model, as well as the uncertainties associated with weather forecasts. In meteorological ensemble forecast systems, the ensemble members often differ in meso- and small-scale features due to sensitivity to initial conditions, but the ensemble mean averages out these variations and is often taken as the operational forecast product for the synoptic weather. We therefore decomposed the uncertainty associated with weather forecasts into (a) the mis-forecast of synoptic weather by the ERA ensemble mean, and (b) variations among the ECMWF ensemble members. Figures 3b and 3c dissect the impacts of these two components of meteorological uncertainty. Assuming that the errors inherited from WRF-GC were similar in all 2DCNN predictions, the ozone hindcast driven by ERA5 meteorological reanalysis (2DCNN-ERA5-2019) would be the best predictions possible for the

2DCNNs (i.e., as close to the observation as possible). Consequently, the deviation of the 2DCNN-SOEF-2019 (driven by actual weather forecast) relative to the 2DCNN-ERA5-2019 hindcasts would be completely due to weather forecast errors. For ozone exceedance cases, the RMSEs of the 2DCNN-SOEF-2019 ensemble mean relative to the 2DCNN-ERA5-2019 hindcast for hourly and MDA8 ozone concentrations were 9.7 ± 1.8 and 10.1 ± 0.7 μg m^{-3}, respectively, at 24-hr lead time (Figure 3b). These purely weather-related RMSEs constituted 38% and 54% of the total ozone forecast RMSEs (Figure 3a) for hourly and MDA8 ozone forecasts at 24-hr lead time, respectively. As such, the errors in synoptic weather forecasts posed a limit to the forecast accuracies of surface ozone at 24-hr lead time and beyond. In comparison, the RMSEs of the 2DCNN-SOEF-2019's ensemble members relative to their ensemble mean were driven by the variations among the weather ensemble forecast members (Figure 3c); these RMSEs were smaller but still sizable, constituting approximately one-third of the total ozone forecast error (Figure 3a).

Figure S9 in Supporting Information S1 further compared the skills of our 2DCNN-SOEF system with the skills of the ECMWF ensemble forecasts for precipitation in Shenzhen. Precipitation, especially the convective precipitation prevailing in the PRD area in summer and fall, is typically regarded as difficult to forecast accurately. In terms of quantitative forecast skills, the RMSEs of precipitation forecasts were large at 24-hr lead time but did not grow significantly with lead time, whereas the RMSEs of high hourly ozone concentration forecasts grew rapidly with lead time (Figure S9a in Supporting Information S1). In terms of categorical forecast skills (Text S6 in Supporting Information S1), the 2DCNN-SOEF was skillful at predicting exceedance events relative both to the climatology (Brier skill score in Figure S9d in Supporting Information S1) and to a random forecast (Receiver Operating Characteristic skill score in Figure S9e in Supporting Information S1) up to 144-hr lead time. The forecast skills of MDA8 ozone exceedances were comparable to those of precipitating events, while the forecast skills of hourly ozone exceedances were considerably lower. If we considered the total hit rates of both exceedance and non-exceedance events (Hiedke skill score in Figure S9f in Supporting Information S1), then the skills of the 2DCNN-SOEFs for both MDA8 and hourly ozone were substantially better than the ECMWF's skills in forecasting precipitating/non-precipitating events.

4. Conclusions

We developed 2-D spatial CNN models to predict surface ozone concentrations from regional weather ensemble forecasts, such that the uncertainty of surface ozone forecasts due to weather forecast uncertainties can be efficiently quantified. Our 2DCNN-SOEF demonstrated comparable performance to current operating forecast systems and met the MEEC's required accuracy for air quality level forecasts up to 144-hr lead time. We showed that uncertainties in weather forecast contributed 38%–54% of the ozone forecast errors at 24-hr lead time and beyond; these weather-related errors cannot be eliminated even with perfect pollutant emission inventories, chemical mechanisms, or numerical techniques. This predictability limit of surface ozone likely varied with geographical locations and seasons and should be accounted for when establishing local air quality forecast accuracy goals and in decision-making for air quality management.

We chose to build our ozone ensemble forecast system using the 2DCNN model not only to accentuate spatial weather patterns, but also to reduce the accumulation of ozone forecast errors with time. Time-series ML/DL models, such as the Long Short Term Memory (LSTM) model and the Recurrent Neural Networks (RNN) have been used in air quality forecasts to capture the temporal evolutions of pollutant concentrations (Athira et al., 2018; Zhao et al., 2019), which are particularly important for short-term forecasts. However, these time-series models tended to accumulate forecast errors over longer lead times (Bi et al., 2022; Freeman et al., 2018). In future work, we intend to combine 2DCNN with LSTM or RNN to improve the 2DCNN's short-term forecast accuracies.

The 2DCNN-SOEF system was computationally efficient for operational applications. On our 48-core Linux server, the WRF-GC model produced a 3-day ozone forecast with only one meteorological realization over the domain of Figure 1 (33,600 grids) in approximately 20 hr of computation time, typical of current regional air quality models. The 2DCNN-SOEF system produced 30 ensemble members of 9-day ozone forecasts, each associated with a different meteorological realization, at only one location in approximately 6 hr (Table S5 in Supporting Information S1). The WRF-GC spent 97% of its computation time on meteorological and chemical calculations. The 2DCNN-SOEF spent 85% of its time downloading the global weather ensemble forecast data, which would not need to be repeated if ozone ensemble forecasts at more locations were desired. On our

hardware, the estimated time for the 2DCNN-SOEF system to generate a 9-day 30-member ensemble hourly ozone forecast for each of the 376 Chinese cities would be 23 hr.

Our work demonstrated the feasibility of developing 2DCNN-SOEF systems for operational air quality forecast and management. As our understanding of ozone photochemistry, regional precursor emissions, and meteorological processes improve, the trainer-induced errors (i.e., errors in the ozone-meteorology relationship simulated by the WRF-GC model) could be further reduced. Our methodology can be extended to develop 2DCNN ensemble forecast models for other air pollutants and other meteorology-dependent environmental risks. Using global weather ensemble forecasts, these risks can be efficiently forecasted at any global location, opening up new applications of deep learning in environmental management.

Data Availability Statement

The source code of the 2-D convolutional neural network-surface ozone ensemble forecast (2DCNN-SOEF) system is available at https://github.com/axzhang1216/2DCNN (last accessed: 13 December 2022). The source code of the Weather Research and Forecasting (WRF)-GC v2.0 model (Feng et al., 2021) is available at https://github.com/jimmielin/wrf-gc-release (last access: 13 December 2022). The WRF-GC and 2DCNN simulation results described in this study are permanently archived at https://doi.org/10.57760/sciencedb.o00005.00025.

References

Athira, V., Geetha, P., Vinayakumar, R., & Soman, K. P. (2018). DeepAirNet: Applying recurrent networks for air quality prediction. *Procedia Computer Science*, *132*, 1394–1403. https://doi.org/10.1016/j.procs.2018.05.068

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., et al. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research*, *106*(D19), 23073–23095. https://doi.org/10.1029/2001JD000807

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2022). Pangu-weather: A 3D high-resolution model for fast and accurate global weather forecast. *arXiv*, 2211. https://doi.org/10.48550/arXiv.2211.02556

Ding, A., Wang, T., Zhao, M., Wang, T., & Li, Z. (2004). Simulation of sea-land breezes and a discussion of their implications on the transport of air pollution during a multi-day ozone episode in the Pearl River Delta of China. *Atmospheric Environment*, *38*(39), 6737–6750. https://doi.org/10.1016/j.atmosenv.2004.09.017

Dong, Y., Li, J., Guo, J., Jiang, Z., Chu, Y., Chang, L., et al. (2020). The impact of synoptic patterns on summertime ozone pollution in the North China Plain. *Science of the Total Environment*, *735*, 139559. https://doi.org/10.1016/j.scitotenv.2020.139559

Ehrendorfer, M. (1997). Predicting the uncertainty of numerical weather forecasts: A review. *Meteorologische Zeitschrift-Berlin*, *6*(4), 147–183. https://doi.org/10.1127/metz/6/1997/147

Feng, X., Lin, H., Fu, T.-M., Sulprizio, M. P., Zhuang, J., Jacob, D. J., et al. (2021). WRF-GC (v2.0): Online two-way coupling of WRF (v3.9.1.1) and GEOS-Chem (v12.7.2) for modeling regional atmospheric chemistry–meteorology interactions. *Geoscientific Model Development*, *14*(6), 3741–3768. https://doi.org/10.5194/gmd-14-3741-2021

Freeman, B. S., Taylor, G., Gharabaghi, B., & Thé, J. (2018). Forecasting air quality time series using deep learning. *Journal of the Air & Waste Management Association*, *68*(8), 866–886. https://doi.org/10.1080/10962247.2018.1459956

Fu, T.-M., & Tian, H. (2019). Climate change penalty to ozone air quality: Review of current understandings and knowledge gaps. *Current Pollution Reports*, *5*(3), 159–171. https://doi.org/10.1007/s40726-019-00115-6

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Huang, L., Liu, S., Yang, Z., Xing, J., Zhang, J., Bian, J., et al. (2021). Exploring deep learning for air pollutant emission estimation. *Geoscientific Model Development*, *14*(7), 4641–4654. https://doi.org/10.5194/gmd-14-4641-2021

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., & Raynaud, L. (2010). *Ensemble of data assimilations at ECMWF*. In *Technical Memorandum 636*. ECMWF. https://doi.org/10.21957/obke4k60

Jacob, D. J., & Winner, D. A. (2009). Effect of climate change on air quality. *Atmospheric Environment*, *43*(1), 51–63. https://doi.org/10.1016/j.atmosenv.2008.09.051

Jiang, Z., Li, J., Lu, X., Gong, C., Zhang, L., & Liao, H. (2021). Impact of Western Pacific subtropical high on ozone pollution over eastern China. *Atmospheric Chemistry and Physics*, *21*(4), 2601–2613. https://doi.org/10.5194/acp-21-2601-2021

Lam, Y. F., Cheung, H. M., & Ying, C. C. (2018). Impact of tropical cyclone track change on regional air quality. *Science of the Total Environment*, *610*, 1347–1355. https://doi.org/10.1016/j.scitotenv.2017.08.100

Li, Y., Zhao, X., Deng, X., & Gao, J. (2022). The impact of peripheral circulation characteristics of typhoon on sustained ozone episodes over the Pearl River Delta region, China. *Atmospheric Chemistry and Physics*, *22*(6), 3861–3873. https://doi.org/10.5194/acp-22-3861-2022

Lin, H., Feng, X., Fu, T.-M., Tian, H., Ma, Y., Zhang, L., et al. (2020). WRF-GC (v1.0): Online coupling of WRF (v3.9.1.1) and GEOS-Chem (v12.2.1) for regional atmospheric chemistry modeling – Part 1: Description of the one-way model. *Geoscientific Model Development*, *13*(7), 3241–3265. https://doi.org/10.5194/gmd-13-3241-2020

Ministry of Ecology and Environment of the People's Republic of China (2020). Technical guideline for numerical forecasting of ambient air quality (HJ 1130-2020). Retrieved from https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/jcffbz/202005/W020200518771113314010.pdf

Ministry of Ecology and Environment of the People's Republic of China (2021). Report on the state of the ecology and environment in China 2021. Retrieved from http://www.gov.cn/xinwen/2022-05/28/5692799/files/349e930e68794f3287888d8dbe9b3ced.pdf

Ministry of Environmental Protection of the People's Republic of China (2015). Emergency management methods for environmental emergencies. Retrieved from https://www.mee.gov.cn/gzk/gz/202112/P020211211513660082369.docx

Molteni, F., Buizza, R., Palmer, T. N., & Petroliagis, T. (1996). The ECMWF ensemble prediction system: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, *122*(529), 73–119. https://doi.org/10.1002/qj.49712252905

Murray, C. J. L., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., et al. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *396*(10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2

NCEP (2000). *NCEP FNL operational model global tropospheric analyses, continuing from July 1999* (Vol. 10). Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6M043C6

Ouyang, S., Deng, T., Liu, R., Chen, J., He, G., Leung, J. C.-H., et al. (2022). Impact of a subtropical high and a typhoon on a severe ozone pollution episode in the Pearl River Delta, China. *Atmospheric Chemistry and Physics*, *22*(16), 10751–10767. https://doi.org/10.5194/acp-22-10751-2022

Palmer, T. N., Barkmeijer, J., Buizza, R., & Petroliagis, T. (1997). The ECMWF ensemble prediction system. *Meteorological Applications*, *4*(4), 301–304. https://doi.org/10.1017/CBO9780511617652.018

Petersen, A. K., Brasseur, G. P., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., et al. (2019). Ensemble forecasts of air quality in eastern China – Part 2: Evaluation of the MarcoPolo–Panda prediction system, version 1. *Geoscientific Model Development*, *12*(3), 1241–1266. https://doi.org/10.5194/gmd-12-1241-2019

Sayeed, A., Choi, Y., Eslami, E., Jung, J., Lops, Y., Salman, A. K., et al. (2021). A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance. *Scientific Reports*, *11*(1), 10891. https://doi.org/10.1038/s41598-021-90446-6

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A., & Jung, J. (2020). Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Networks*, *121*, 396–408. https://doi.org/10.1016/j.neunet.2019.09.033

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D., Duda, M. G., et al. (2008). *A description of the advanced research WRF version 3*. NCAR/TN-475+ STR, University Corporation for Atmospheric Research. National Center for Atmospheric Research Boulder. https://doi.org/10.5065/D68S4MVH

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., et al. (2019). *A description of the advanced research WRF model version 4* (Vol. 145, p. 145). NCAR/TN-556+STR National Center for Atmospheric Research. https://doi.org/10.5065/1dfh-6p97

Slingo, J., & Palmer, T. (2011). Uncertainty in weather and climate prediction. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *369*(1956), 4751–4767. https://doi.org/10.1098/rsta.2011.0161

Sun, H., Shin, Y. M., Xia, M., Ke, S., Wan, M., Yuan, L., et al. (2021). Spatial resolved surface ozone with urban and rural differentiation during 1990–2019: A space–time Bayesian neural network downscaler. *Environmental Science & Technology*, *56*(11), 7337–7349. https://doi.org/10.1021/acs.est.1c04797

Sun, Z., & Archibald, A. T. (2021). Multi-stage ensemble-learning-based model fusion for surface ozone simulations: A focus on CMIP6 models. *Environmental Science and Ecotechnology*, *8*, 100124. https://doi.org/10.1016/j.ese.2021.100124

U. S. Environmental Protection Agency. (2015). Air quality guide for ozone. Retrieved from https://www.epa.gov/sites/production/files/2017-12/documents/air-quality-guide_ozone_2015.pdf

U.S. Environmental Protection Agency. (2021). Our Nation's air: Trends through 2021. Retrieved from https://gispub.epa.gov/air/trendsreport/2022

Wang, N., Huang, X., Xu, J., Wang, T., Tan, Z., & Ding, A. (2022). Typhoon-boosted biogenic emission aggravates cross-regional ozone pollution in China. *Science Advances*, *8*(2), eabl6166. https://doi.org/10.1126/sciadv.abl6166

Wang, X., Fu, T.-M., Zhang, L., Cao, H., Zhang, Q., Ma, H., et al. (2021). Sensitivies of ozone air pollution in the Beijing–Tianjin–Hebei area to local and upwind precursor emissions using adjoint modeling. *Environmental Science & Technology*, *55*(9), 5752–5762. https://doi.org/10.1021/acs.est.1c00131

Weng, X., Forster, G. L., & Nowack, P. (2022). A machine learning approach to quantify meteorological drivers of ozone pollution in China from 2015 to 2019. *Atmospheric Chemistry and Physics*, *22*(12), 8385–8402. https://doi.org/10.5194/acp-22-8385-2022

World Health Organization (2021). WHO global air quality guidelines. Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. In *Executive summary*. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/345334/9789240034433-eng.pdf

Xing, J., Zheng, S., Ding, D., Kelly, J. T., Wang, S., Li, S., et al. (2020). Deep learning for prediction of the air quality response to emission changes. *Environmental Science & Technology*, *54*(14), 8589–8600. https://doi.org/10.1021/acs.est.0c02923

Zhang, F., Bei, N., Nielsen-Gammon, J. W., Li, G., Zhang, R., Stuart, A., & Aksoy, A. (2007). Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. *Journal of Geophysical Research*, *112*(D4), D04304. https://doi.org/10.1029/2006JD007429

Zhao, J., Deng, F., Cai, Y., & Chen, J. (2019). Long short-term memory-fully connected (Lstm-fc) neural network for $PM_{2.5}$ concentration prediction. *Chemosphere*, *220*, 486–492. https://doi.org/10.1016/j.chemosphere.2018.12.128

References From the Supporting Information

Bates, K. H., & Jacob, D. J. (2019). A new model mechanism for atmospheric oxidation of isoprene: Global effects on oxidants, nitrogen oxides, organic products, and secondary organic aerosol. *Atmospheric Chemistry and Physics*, *19*(14), 9613–9640. https://doi.org/10.5194/acp-19-9613-2019

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Chen, Z., Li, R., Chen, D., Zhuang, Y., Gao, B., Yang, L., & Li, M. (2020). Understanding the causal influence of major meteorological factors on ground ozone concentrations across China. *Journal of Cleaner Production*, *242*, 118498. https://doi.org/10.1016/j.jclepro.2019.118498

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., et al. (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research*, *101*(D3), 7251–7268. https://doi.org/10.1029/95JD02165

China National Environmental Monitoring Centre. (2017). *Technical guide to ambient air quality forecasting and warning methodology* (2nd ed.). China Environmental Science Press.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., et al. (2010). Description and evaluation of the model for ozone and related chemical tracers, version 4 (MOZART-4). *Geoscientific Model Development*, *3*(1), 43–67. https://doi.org/10.5194/gmd-3-43-2010

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., & Wang, X. (2012). The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. *Geoscientific Model Development*, *5*(6), 1471–1492. https://doi.org/10.5194/gmd-5-1471-2012

Hamill, T. M., & Juras, J. (2006). Measuring forecast skill: Is it real skill or is it the varying climatology? *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography*, *132*(621C), 2905–2923. https://doi.org/10.1256/qj.06.25

Heald, C. L., Ridley, D. A., Kroll, J. H., Barrett, S. R. H., Cady-Pereira, K. E., Alvarado, M. J., & Holmes, C. D. (2014). Contrasting the direct radiative effect and direct radiative forcing of aerosols. *Atmospheric Chemistry and Physics*, *14*(11), 5513–5527. https://doi.org/10.5194/acp-14-5513-2014

Heidke, P. (1926). Berechnung des erfolges und der güte der windstärkevorhersagen im sturmwarnungsdienst. *Geografiska Annaler*, *8*(4), 301–349. https://doi.org/10.1080/20014422.1926.11881138

Hudman, R. C., Russell, A. R., Valin, L. C., & Cohen, R. C. (2010). Interannual variability in soil nitric oxide emissions over the United States as viewed from space. *Atmospheric Chemistry and Physics*, *10*(20), 9943–9952. https://doi.org/10.5194/acp-10-9943-2010

Hyvärinen, O. (2014). A probabilistic derivation of Heidke skill score. *Weather and Forecasting*, *29*(1), 177–181. https://doi.org/10.1175/WAF-D-13-00103.1

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*(D13), D13103. https://doi.org/10.1029/2008JD009944

Kharin, V. V., & Zwiers, F. W. (2003). On the ROC score of probability forecasts. *Journal of Climate*, *16*(24), 4145–4150. https://doi.org/10.1175/1520-0442(2003)016<4145:OTRSOP>2.0.CO;2

Legg, T. P., & Mylne, K. R. (2004). Early warnings of severe weather from ensemble forecast information. *Weather and Forecasting*, *19*(5), 891–906. https://doi.org/10.1175/1520-0434(2004)019<0891:EWOSWF>2.0.CO;2

Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., & Bates, K. H. (2019). Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. *Proceedings of the National Academy of Sciences*, *116*(2), 422–427. https://doi.org/10.1073/pnas.1812168116

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., et al. (2017). Anthropogenic emission inventories in China: A review. *National Science Review*, *4*(6), 834–866. https://doi.org/10.1093/nsr/nwx150

Lu, K., Zhang, Y., Su, H., Shao, M., Zeng, L., Zhong, L., et al. (2010). Regional ozone pollution and key controlling factors of photochemical ozone production in Pearl River Delta during summer time. *Scientia Sinica Chimica*, *40*(4), 407–420. https://doi.org/10.1007/s11426-010-0055-6

Mellor, G. L., & Yamada, T. (1982). Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, *20*(4), 851–875. https://doi.org/10.1029/RG020i004p00851

Ministry of Ecology and Environment of the People's Republic of China (2020b). Technical guideline for numerical forecasting of ambient air quality (HJ 1130-2020). Retrieved from https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/jcffbz/202005/W020200518771113314010.pdf

Ministry of Environmental Protection of the People's Republic of China (2012). Ambient air quality standard (GB 3095-2012). Retrieved from https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/dqhjbh/dqhjzlbz/201203/W020120410330232398521.pdf

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16663–16682. https://doi.org/10.1029/97JD00237

Monin, A. S., & Obukhov, A. M. (1954). Basic laws of turbulent mixing in the ground surface layer. *Trudy - Geologicheskiy Institut, Akademiya Nauk SSSR*, *151*, 163–187. https://mcnaughty.com/keith/papers/Monin_and_Obukhov_1954.pdf

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly Weather Review*, *137*(3), 991–1007. https://doi.org/10.1175/2008MWR2556.1

Murray, L. T., Jacob, D. J., Logan, J. A., Hudman, R. C., & Koshak, W. J. (2012). Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data. *Journal of Geophysical Research*, *117*(D20). https://doi.org/10.1029/2012JD017934

Nakanishi, M., & Niino, H. (2009). Development of an improved turbulence closure model for the atmospheric boundary layer. *Journal of the Meteorological Society of Japan. Ser. II*, *87*(5), 895–912. https://doi.org/10.2151/jmsj.87.895

Pound, R. J., Sherwen, T., Helmig, D., Carpenter, L. J., & Evans, M. J. (2020). Influences of oceanic ozone deposition on tropospheric photochemistry. *Atmospheric Chemistry and Physics*, *20*(7), 4227–4239. https://doi.org/10.5194/acp-20-4227-2020

Richards, F., & Arkin, P. (1981). On the relationship between satellite-observed cloud cover and precipitation. *Monthly Weather Review*, *109*(5), 1081–1093. https://doi.org/10.1175/1520-0493(1981)109<1081:OTRBSO>2.0.CO;2

Safieddine, S. A., & Heald, C. L. (2017). A global assessment of dissolved organic carbon in precipitation. *Geophysical Research Letters*, *44*(22). https://doi.org/10.1002/2017GL075270

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, *117*(8), 1779–1800. https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2

Stauffer, D. R., & Seaman, N. L. (1990). Use of four-dimensional data assimilation in a limited-area mesoscale model. Part i: Experiments with synoptic-scale data. *Monthly Weather Review*, *118*(6), 1250–1277. https://doi.org/10.1175/1520-0493(1990)118<1250:UOFDDA>2.0.CO;2

Weigel, A. P., Liniger, M. A., & Appenzeller, C. (2007). The discrete Brier and ranked probability skill scores. *Monthly Weather Review*, *135*(1), 118–124. https://doi.org/10.1175/MWR3280.1

Zhang, C., Wang, Y., & Hamilton, K. (2011). Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme. *Monthly Weather Review*, *139*(11), 3489–3513. https://doi.org/10.1175/MWR-D-10-05091.1

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., et al. (2018). Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmospheric Chemistry and Physics*, *18*(19), 14095–14111. https://doi.org/10.5194/acp-18-14095-2018

Zheng, B., Zhang, Q., Geng, G., Chen, C., Shi, Q., Cui, M., et al. (2021). Changes in China's anthropogenic emissions and air quality during the COVID-19 pandemic in 2020. *Earth System Science Data*, *13*(6), 2895–2907. https://doi.org/10.5194/essd-13-2895-2021