GeoHealth

RESEARCH ARTICLE
10.1029/2021GH000570

Special Section:
Atmospheric PM2.5 in China: indoor, outdoor, and health effects

This article is a companion to Conibear et al. (2021), https://doi.org/10.1029/2021GH000391; and Conibear et al. (2022a), https://doi.org/10.1029/2021GH000567.

Key Points:

- We developed emulators to predict air quality and chronic health impacts across China from changes in emissions (based on 2015 data)
- Annual−mean $PM_{2.5}$ exposure and the associated disease burden were most sensitive to changes in industrial and residential emissions
- Removing emissions from five key sectors in China does not attain the World Health Organization Annual Guideline due to remaining pollution from other sources

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:
L. Conibear,
laconibear@gmail.com

Citation:
Conibear, L., Reddington, C. L., Silver, B. J., Chen, Y., Knote, C., Arnold, S. R., & Spracklen, D. V. (2022). Sensitivity of air pollution exposure and disease burden to emission changes in China using machine learning emulation. *GeoHealth*, *6*, e2021GH000570. https://doi.org/10.1029/2021GH000570

Received 1 FEB 2022
Accepted 16 MAY 2022

Sensitivity of Air Pollution Exposure and Disease Burden to Emission Changes in China Using Machine Learning Emulation

Luke Conibear[1], Carly L. Reddington[1], Ben J. Silver[1], Ying Chen[2], Christoph Knote[3], Stephen R. Arnold[1], and Dominick V. Spracklen[1]

[1]School of Earth and Environment, Institute for Climate and Atmospheric Science, University of Leeds, Leeds, UK, [2]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK, [3]Faculty of Medicine, University of Augsburg, Augsburg, Germany

Abstract Machine learning models can emulate chemical transport models, reducing computational costs and enabling more experimentation. We developed emulators to predict annual−mean fine particulate matter ($PM_{2.5}$) and ozone (O_3) concentrations and their associated chronic health impacts from changes in five major emission sectors (residential, industrial, land transport, agriculture, and power generation) in China. The emulators predicted 99.9% of the variance in $PM_{2.5}$ and O_3 concentrations. We used these emulators to estimate how emission reductions can attain air quality targets. In 2015, we estimate that $PM_{2.5}$ exposure was 47.4 μg m^{-3} and O_3 exposure was 43.8 ppb, associated with 2,189,700 (95% uncertainty interval, 95UI: 1,948,000–2,427,300) premature deaths per year, primarily from $PM_{2.5}$ exposure (98%). $PM_{2.5}$ exposure and the associated disease burden were most sensitive to industry and residential emissions. We explore the sensitivity of exposure and health to different combinations of emission reductions. The National Air Quality Target (35 μg m^{-3}) for $PM_{2.5}$ concentrations can be attained nationally with emission reductions of 72% in industrial, 57% in residential, 36% in land transport, 35% in agricultural, and 33% in power generation emissions. We show that complete removal of emissions from these five sectors does not enable the attainment of the WHO Annual Guideline (5 μg m^{-3}) due to remaining air pollution from other sources. Our work provides the first assessment of how air pollution exposure and disease burden in China varies as emissions change across these five sectors and highlights the value of emulators in air quality research.

Plain Language Summary The ability of air quality models to help address important public health problems is limited by their high computational costs. Machine learning models can help by accurately representing these complicated air quality models for specific prediction tasks. These machine learning models can then be run many times at a fraction of the time and cost. Here, we developed machine learning models to predict long–term air quality and the associated health impacts in China from changes in emissions. We found that reducing emissions linearly improves particulate air quality and public health. The fractional improvements in public health were smaller than the fractional improvements in air quality. Removing emissions from five key sectors (residential, industrial, land transport, agriculture, and power generation) does not attain the World Health Organization Annual Guideline because of remaining pollution from other sources, such as from alternative anthropogenic emissions inside China, anthropogenic emissions outside China, and natural emissions. This work illustrates the broad reach of emulators in air pollution research.

1. Introduction

Air pollution exposure is a key public health problem in China (GBD, 2019 Risk Factors Collaborators, 2020; Yin et al., 2020). In recent years, particulate air quality has improved, primarily attributed to reductions in anthropogenic emissions (Cheng et al., 2019; Ding et al., 2019; Li et al., 2020, 2019; Silver, Conibear, et al., 2020; Silver, He et al., 2020; Silver et al., 2018; Zhai et al., 2019). However, the loss of healthy life from air pollution exposure remains substantial (Conibear, Reddington, Silver, Knote, et al., 2021; Silver, Conibear, et al., 2020; Zhao et al., 2018) and further emission reductions are required to improve air quality.

Numerical chemical transport models (CTMs) are useful for simulating air quality and its driving processes. CTMs discretize the atmosphere into cells on a three-dimensional grid and compute complex processes, mechanisms, laws, and parameterisations at high temporal resolutions on this grid (Brasseur & Jacob, 2017). However,

Author Contributions:
Conceptualization: Luke Conibear, Stephen R. Arnold, Dominick V. Spracklen
Data curation: Luke Conibear, Carly L. Reddington, Ben J. Silver
Formal analysis: Luke Conibear
Funding acquisition: Stephen R. Arnold, Dominick V. Spracklen
Investigation: Luke Conibear
Methodology: Luke Conibear, Ying Chen
Project Administration: Stephen R. Arnold, Dominick V. Spracklen
Resources: Stephen R. Arnold, Dominick V. Spracklen
Software: Luke Conibear, Christoph Knote
Validation: Luke Conibear
Visualization: Luke Conibear, Dominick V. Spracklen
Writing – original draft: Luke Conibear
Writing – review & editing: Luke Conibear, Carly L. Reddington, Ying Chen, Stephen R. Arnold

these complex CTMs have high computational costs. To reduce costs, some approaches are to reduce the model complexity, reduce the model resolution or precision (Palmer, 2015), reduce the number of experiments, or to use simplified models. To reduce computational demand, a wide range of reduced−complexity or reduced−form air quality models have been developed to simplify these complex processes (Buonocore et al., 2014; Carnevale et al., 2009; Foley et al., 2014; Heo, Adams & Gao, 2016a, 2016b; Henze et al., 2007; Seinfeld & Pandis, 2016; Tessum et al., 2017). For example, InMAP (Intervention Model for Air Pollution) is a reduced−form air quality model that decreases computational costs via simplified representations of atmospheric processes (Tessum et al., 2017). In contrast, the emulators are statistical machine learning models that decrease computational costs via mapping specific associations. Emulators learn these specific input−output associations from full CTM simulations. These emulators are often designed using Gaussian process regressors (O'Hagan, 2006; Rasmussen & Williams, 2006), due to their flexibility, accuracy, and skill with smaller data sets. Emulators are computationally expensive to build as their training data requires many CTM runs, though they are substantially cheaper to run once built enabling many more experiments to be explored.

Previous studies have used emulators for prediction of air quality (Beddows et al., 2017; Chen et al., 2020; Conibear, Reddington, Silver, Chen, et al., 2021), weather (Chantry, Christensen, et al., 2021; Gettelman et al., 2021; Weyn et al., 2019), and climate (Beusch et al., 2020; Chantry, Christensen, et al., 2021; Holden et al., 2019; Ott et al., 2020; Scher, 2018; Tran et al., 2016). For example, the air quality prediction studies used emulators to analyze the drivers of an ozone (O_3) pollution episode in the United Kingdom (Beddows et al., 2017) and emission reduction strategies in India (Chen et al., 2020). Some studies have used machine learning models to represent processes, such as radiation (Chevallier et al., 2000; Krasnopolsky et al., 2005), convection (Beucler et al., 2020; Brenowitz & Bretherton, 2018; Gentine et al., 2018; Han et al., 2020; O'Gorman & Dwyer, 2018; Rasp et al., 2018; Yuval & O'Gorman, 2020), chemistry (Ivatt & Evans, 2020; Keller & Evans, 2019; Kelp et al., 2020; Kelp et al., 2022), physics (Chantry, Hatfield et al., 2021; Harder et al., 2021; Hatfield et al., 2021; Hughes et al., 2018; Krasnopolsky, 2020; Silva et al., 2021), and land surface models (Dagon et al., 2020). Many studies have used emulators to explore uncertainties and sensitivities (Aleksankina et al., 2019; Carslaw et al., 2013; Chang et al., 2016; Johnson et al., 2015; Lee et al., 2012; Lee et al., 2011; Lee et al., 2016; McCoy et al., 2020; Nicely et al., 2020; Ryan & Wild, 2021; Ryan et al., 2018; Rybarczyk & Zalakeviciute, 2018; Salter et al., 2018; Watson-Parris, 2021; Watson-Parris et al., 2020; Watson-Parris et al., 2021; Wild et al., 2020).

In our previous work, we developed emulators to predict winter (January 2015) ambient fine particulate matter ($PM_{2.5}$) concentrations from emission changes across China (Conibear, Reddington, Silver, Chen, et al., 2021). Here, we further develop these emulators for annual exposure (2015) to multiple air pollutants ($PM_{2.5}$ and O_3) and to assess the chronic health impacts. To our knowledge, this is the first study using emulators to predict long−term (annual) air quality and the public health benefits attributed to different emission control strategies in China.

2. Methods

2.1. Simulator

Simulations were conducted using WRFChem (Weather Research and Forecasting model online–coupled with Chemistry) version 3.7.1 (Grell et al., 2005; Skamarock et al., 2008). Each simulation was for the whole of 2015 with one–month spin–up over China at 30 km (0.3°) horizontal resolution. There were 50 simulations for the training data and five additional simulations for the test data. The simulations differed only in the scaling of the anthropogenic emissions over China, determined from separate maxi−min Latin hypercube space–filling designs (Tables S1 and S2 in Supporting Information S1). The version of WRFChem used here was described and evaluated in our previous work (Conibear, Reddington, Silver, Chen, et al., 2021; Conibear, Reddington, Silver, Knote, et al., 2021; Reddington et al., 2019; Silver, Conibear, et al., 2020).

Sectoral and speciated anthropogenic emissions inside China were from the MEIC (Multi–resolution Emission Inventory for China) emission inventory for 2015 at 0.25° × 0.25° horizontal resolution (Li, Liu, et al., 2017; Li, Zhang, et al., 2017; MEIC Research Group & Tsinghua University, 2019; Zheng et al., 2018). Sectoral and speciated anthropogenic emissions outside China were from EDGAR−HTAP (Emission Database for Global Atmospheric Research with Task Force on Hemispheric Transport of Air Pollution) version 2.2 for 2010 at 0.1° × 0.1° horizontal resolution (Janssens-Maenhout et al., 2015). Anthropogenic emissions are largest over East, North,

and South China, as well as across South Asia (Figures S1−S4 in Supporting Information S1). A diurnal cycle was applied to the anthropogenic emissions (Qi et al., 2017; Zheng et al., 2017).

Gas phase chemistry was simulated using the extended MOZART (Model for Ozone and Related Chemical Tracers) scheme (Emmons et al., 2010; Hodzic & Jimenez, 2011; Knote et al., 2014). Aerosol physics and chemistry was simulated using the updated MOSAIC (Model for Simulating Aerosol Interactions and Chemistry) scheme with aqueous chemistry (Alma Hodzic & Knote, 2014; Zaveri et al., 2008). Secondary organic aerosol (SOA) formation was based on an updated volatility basis set mechanism (Knote et al., 2015).

We evaluated the simulator (WRFChem) against $PM_{2.5}$ and O_3 measurements from 1,633 sites (Jin et al., 2020; Silver et al., 2018). The normalized mean bias factor (NMBF) and the normalized mean absolute error factor (NMAEF) were used to evaluate the simulator (Yu et al., 2006). For example, a NMBF of −0.5 means that the simulator underestimated observations by 50% on average, and a NMAEF of 0.5 means that the simulator had an absolute gross error of 0.5 times the mean observation. Here, the simulator underestimated annual−mean $PM_{2.5}$ concentrations (NMBF = −0.05 and NMAEF = 0.18) and overestimated maximum 6−monthly−mean daily−maximum 8−hour (6mDM8h) O_3 concentrations (NMBF = 0.39 and NMAEF = 0.40) across China (Figure S5 in Supporting Information S1). To provide the closest match with observations, we tuned the $PM_{2.5}$ and O_3 concentrations. Tuning was completed by scaling the model to match observations by prefecture if measurements were available, otherwise scalings were applied by province (administrative division). The tuned model had reduced bias and error for both annual−mean $PM_{2.5}$ concentrations (NMBF = 0.02 and NMAEF = 0.10) and 6mDM8h O_3 concentrations (NMBF = 0.03 and NMAEF = 0.11) across China. The scaling allowed us to accurately predict the spatial pattern and magnitude of $PM_{2.5}$ and O_3 concentrations across China.

2.2. Emulator

We developed emulators to make computationally efficient predictions of the WRFChem model, as described in our previous work (Conibear, Reddington, Silver, Chen, et al., 2021). Here, the emulator approach from our previous work was extended from January 2015 to the year of 2015, for O_3 concentrations, and the air pollution disease burden. The emulator workflow is summarized in Figure S6 in Supporting Information S1.

The emulators predict air quality across China from fractional changes in anthropogenic emissions. The emulator inputs were anthropogenic emissions from the residential (RES), industrial (IND), land transport (TRA), agricultural (AGR), and power generation (ENE) sectors. For the emulator inputs, all species from each sector were scaled between 0%–150%. These emulator inputs were from simulator data of 50 training runs and 5 test runs, based on separate maxi−min Latin hypercube space–filling designs. The training data was designed to cover all of the input distributions. The emulators were trained on the raw simulation data, before the control scaling factors were applied for tuning.

Individual emulators were developed for each output of annual−mean $PM_{2.5}$ concentrations and 6mDM8h O_3 concentrations, and each WRFChem grid cell across China (15,278 grid cells in total) to capture the spatial distribution of each pollutant. These outputs were chosen as they are the metrics used in the health impact assessment. This meant we developed a total of 30,556 separate emulators. The emulators are based on annual average values and do not have information of finer time scales.

The optimized emulator designs included an input preprocessor (Yeo & Johnson, 2000) and a Gaussian process regressor with a Matern 5/2 kernel (Conibear, 2021). Gaussian process regressors update a prior function over the inputs to a posterior function including observations (i.e., the training data) (Rasmussen & Williams, 2006). Bayesian inference is then used to sample from this posterior function to produce the Gaussian output. Gaussian process regressors notice trends well when similar inputs have similar outputs, and are flexible and accurate with smaller data sets. Our emulators were based on Gaussian process regressors as they are accurate with only a few training samples. This was a key limitation as each training sample is an annual CTM simulation. Other emulator design options, such as deep neural networks, often require much larger training datasets to avoid overfitting the limited data (Watson−Parris, 2021). Recent developments in machine learning architectures may overcome this limitation to improve emulator accuracy and scope, for example, with deep neural architecture search (Kasim et al., 2022).

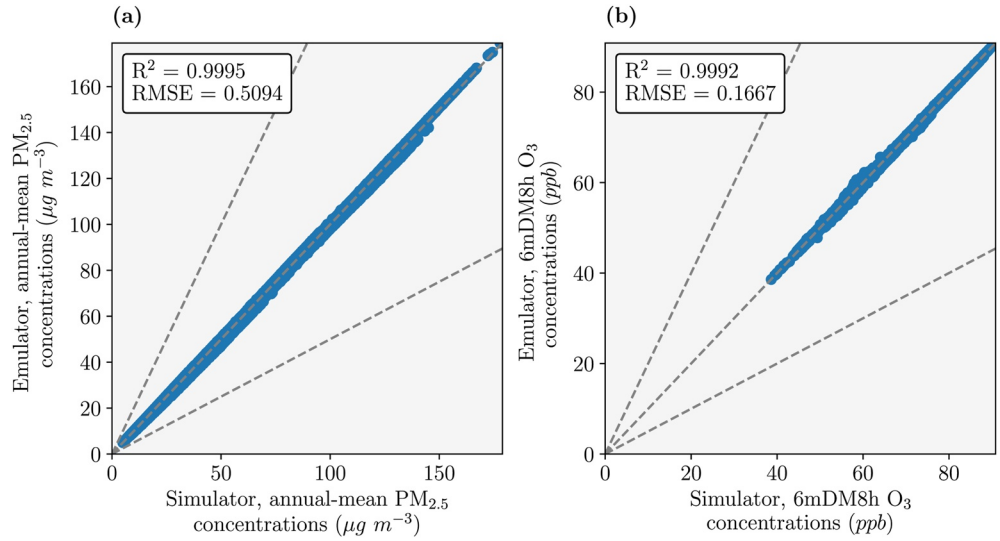

Figure 1. Evaluation of emulators on the unseen test data for concentrations of (a) fine particulate matter ($PM_{2.5}$, annual−mean) and (b) ozone (O_3, maximum 6−monthly−mean daily−maximum 8−hour, 6mDM8h). Evaluation metrics used were the coefficient of determination (R^2) and the root mean squared error (RMSE).

The creation of the simulation data (i.e., 55 annual WRFChem runs) was computationally expensive, using 320 CPUs (Central Processing Units, Intel Xeon Gold 6,138) for approximately 1 year of wall time. The training of the emulators used 150 CPUs for approximately 1 hr. Using the emulators per prediction run on the order of seconds on 1 CPU. Hence, the key bottleneck is simulating the atmosphere using complex numerical models. Reducing the computational burden of this step is an important area of future research.

The emulators were specific to the data they were trained on. The emulators make predictions based on associational knowledge, rather than explanatory knowledge (Deutsch, 2012; Pearl, 2019). The emulators were used to predict air quality concentrations for all emission configurations within a 0%–150% matrix of emission scaling factors at 20% increments (32,768 emission configurations).

Figure 1 shows the evaluation of the emulators on the unseen test data. For $PM_{2.5}$ concentrations, the emulators have a coefficient of determination (R^2) value of 0.9995 and a root mean squared error (RMSE) value of 0.5094 μg m^{-3}. For O_3 concentrations, the emulators have an R^2 value of 0.9992 and a RMSE value of 0.1667 ppb. This means that the emulators can accurately predict 99.9% of the variance in both $PM_{2.5}$ and O_3 concentrations for any similar emission configuration.

2.3. Health Impact Assessment

The health impact assessment estimated the disease burden attributable to $PM_{2.5}$ and O_3 exposure using population attributable fractions (PAF) of relative risk (RR). Exposure variations were used to predict associated outcome variations.

The exposure to annual−mean $PM_{2.5}$ (z) per grid cell was relative to the counterfactual exposure level of 2.4 μg m^{-3} (cf.) where no excess risk was assumed (Equation 1). The RR for a specific exposure and population age group was estimated through the GEMM (Global Exposure Mortality Model) (Burnett et al., 2018). The RR was a function of the parameters θ, α, μ, and ν (Equation 2) as defined in Supplementary Table 2 of Conibear, Reddington, Silver, Knote, et al. (2021). We used the GEMM for non−accidental mortality (non−communicable disease, NCD, plus lower respiratory infections, LRI), using parameters that included the China cohort, with age−specific modifiers for adults over 25 years of age in five−year intervals. The GEMM functions have mean, lower, and upper uncertainty intervals. The PAF was estimated as a function of the RR and the population count (P, Equation 3).

$$z = \max(0, PM_{2.5} - cf) \quad (1)$$

$$\mathrm{RR}(z,\mathrm{age}) = e^{\left\{\theta \frac{\log(1+\frac{z}{\alpha})}{1+e^{(\frac{\mu-z}{\nu})}}\right\}} \quad (2)$$

$$\mathrm{PAF} = P \times \left(1 - \frac{1}{\mathrm{RR}(z,\mathrm{age})}\right) \quad (3)$$

The health impact assessment for O_3 exposure followed the methodology of the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) for 2017 (GBD, 2017 Risk Factor Collaborators, 2018). The exposure to O_3 (z) per grid cell was calculated as the change in 6mDM8h O_3 concentrations relative to the counterfactual exposure level of 35.7 ppb (cf.) where no excess risk was assumed (Equation 4) (Turner et al., 2016). The 6mDM8h metric was calculated by quantifying 24 separate 8−hour rolling mean O_3 concentrations, finding the maximum of these each day, creating 12 separate 6−monthly means to account for seasonal variations, and finding the maximum of these over the year. The PAF was a function of the hazard ratio (HR), which was 1.06 (95UI: 1.02–1.10) for chronic obstructive pulmonary disease (COPD), based on data from five epidemiological cohorts (Equation 5) (GBD, 2017 Risk Factor Collaborators, 2018).

$$\mathrm{z} = \max(0, \mathrm{O_3} - \mathrm{cf}) \quad (4)$$

$$\mathrm{PAF} = P \times \left(1 - e^{z \frac{\log HR}{10}}\right) \quad (5)$$

Premature mortality (MORT), years of life lost (YLL), and years lived with disability (YLD) per exposure, health outcome, age bracket, and grid cell were estimated as a function of the PAF and the corresponding baseline mortality and morbidity rate (I_{MORT}, I_{YLL}, and I_{YLD}) following Equations 6–8, respectively. Disability−adjusted life years (DALYs) were estimated as the total of YLL and YLD (Equation 9). The rates of MORT, YLL, YLD, and DALYs were calculated per 100,000 people.

$$\mathrm{MORT} = \mathrm{PAF} \times \mathrm{I_{MORT}} \quad (6)$$

$$\mathrm{YLL} = \mathrm{PAF} \times \mathrm{I_{YLL}} \quad (7)$$

$$\mathrm{YLD} = \mathrm{PAF} \times \mathrm{I_{YLD}} \quad (8)$$

$$\mathrm{DALYs} = \mathrm{YLL} + \mathrm{YLD} \quad (9)$$

The United Nations adjusted population count data set for 2015 at 0.25° × 0.25° resolution was obtained from the Gridded Population of the World, Version 4 (Center for International Earth Science Information Network & NASA Socioeconomic Data and Applications Center, 2016). Population age composition was taken from the GBD2017 for 2015 for adults of 25–80 years of age in 5−year intervals, and for 80 years plus (Global Burden of Disease Study, 2017, 2018). Cause−specific (NCD, LRI, and COPD) baseline mortality and morbidity rates were taken from the GBD2017 for 2015 for MORT, YLL, and YLD for each age bracket (Institute for Health Metrics and Evaluation, 2020).

Shapefiles were used to aggregate results at the country, province, and prefecture level (Hijmans et al., 2020). Regional groupings were applied as follows: North China (Beijing, Tianjin, Hebei, Shanxi, and Inner Mongolia), North East China (Liaoning, Jilin, and Heilongjiang), East China (Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, and Shandong), South Central China (Henan, Hubei, Hunan, Guangdong, Guangxi, Hainan, Hong Kong, and Macau) including the Guangdong−Hong Kong−Macau Greater Bay Area (GBA), South West China (Chongqing, Sichuan, Guizhou, Yunnan, and Tibet), and North West China (Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang), and the GBA individually.

Uncertainty intervals at the 95% confidence level were estimated using the derived uncertainty intervals from the exposure−outcome associations, baseline mortality and morbidity rates, and population age fractions. Health impact assessments of the disease burden associated with air pollution exposure have many uncertainties (Nethery & Dominici, 2019).

Table 1
The Emulated Baseline (2015) Exposure and Disease Burden for Fine Particulate Matter ($PM_{2.5}$, Annual−Mean) and Ozone (O_3, Maximum 6−Monthly−Mean Daily−Maximum 8−Hour, 6mDM8h) Across Regions in China

Baseline (2015)	China	GBA	North China	North east China	East China	South central China	South west China	North west China
Annual− mean $PM_{2.5}$ exposure (μg m^{-3})	47.4	26.8	62.3	42.4	47.0	45.3	29.4	37.6
6mDM8h O_3 exposure (ppb)	43.8	35.7	48.5	41.8	42.8	41.9	41.5	46.2
MORT $PM_{2.5}$ (deaths yr^{-1})	2,143,700 (1,916,200–2,382,400)	66,600 (59,300–74,300)	335,700 (300,800–372,200)	158,200 (141,300–175,900)	651,100 (582,300–723,200)	593,600 (530,600–659,600)	258,200 (230,200–287,800)	147,000 (131,100–163,700)
MORT O_3 (deaths yr^{-1})	46,000 (31,800–64,900)	1,300 (900–1,900)	8,300 (5,800–11,700)	3,300 (2,300–4,700)	15,300 (10,600–21,500)	10,900 (7,500–15,300)	4,900 (3,400–7,000)	3,200 (2,200–4,600)
DALYs rate $PM_{2.5}$ (DALYs 100,000 $people^{-1}$ yr^{-1})	4,219 (3,412–5,178)	4,318 (3,490–5,304)	5,651 (4,591–6,906)	4,768 (3,861–5,846)	5,431 (4,407–6,645)	5,330 (4,323–6,524)	3,574 (2,885–4,396)	4,231 (3,420–5,197)
DALYs rate O_3 (DALYs 100,000 $people^{-1}$ yr^{-1})	68 (44–99)	51 (33–74)	112 (73–161)	70 (45–101)	72 (47–104)	64 (41–93)	40 (26–58)	84 (54–122)

Note. The disease burden is given by the annual number of premature mortalities (MORT) and the annual rate of disability−adjusted life years (DALYs) per 100,000 people.

3. Results and Discussion

In the results and discussion, $PM_{2.5}$ concentrations are ambient annual−means and O_3 concentrations are ambient 6mDM8h. Exposures are population−weighted concentrations. Air quality standards for O_3 concentrations in units of μg m^{-3} are converted to ppb using the conversion factor of one ppb being approximately equal to 2 μg m^{-3} (Fleming et al., 2018).

3.1. Emulated Baseline $PM_{2.5}$ and O_3 Exposure and Disease Burden for 2015

Baseline (i.e., for 2015 with all emission sectors at 100%) $PM_{2.5}$ exposure in China is 47.4 μg m^{-3}, with higher exposure over North, East, and South Central China, and lower exposure in the GBA, South West, and North West China (Table 1 and Figure S5 in Supporting Information S1). $PM_{2.5}$ exposure is highest across the provinces of Hebei, Beijing, Shandong, Henan, and Anhui in North and East China. The National Air Quality Target (35 μg m^{-3}) is achieved in GBA and South West China. The baseline annual disease burden associated with $PM_{2.5}$ exposure in China is estimated as 2,143,700 (95UI: 1,916,200−2,382,400) premature mortalities and 4,219 (95UI: 3,412−5,178) DALYs per 100,000 people. Our estimated disease burdens for 2015 are similar to the previous estimate of 2,440,000 (95UI: 2,046,600−2,808,300) premature mortalities from Burnett et al. (2018). The disease burden rate is largest in North, East, and South Central China.

Baseline O_3 exposure in China is 43.8 ppb, with higher exposure over North, North West, and East China, and lower exposure in the GBA, North East, and South Central China (Table 1 and Figure S5 in Supporting Information S1). O_3 exposure is highest across the provinces of Hubei, Beijing, and Shandong. The National Air Quality Target (80 ppb) and the World Health Organization (WHO) guideline (50 ppb) are achieved in all regions at the baseline (World Health Organization, 2021). The baseline annual disease burden associated with O_3 exposure in China is estimated as 46,000 (95UI: 31,800−64,900) premature mortalities and 68 (95UI: 44–99) DALYs per 100,000 people. The disease burden rate is largest in North, North West, and East China.

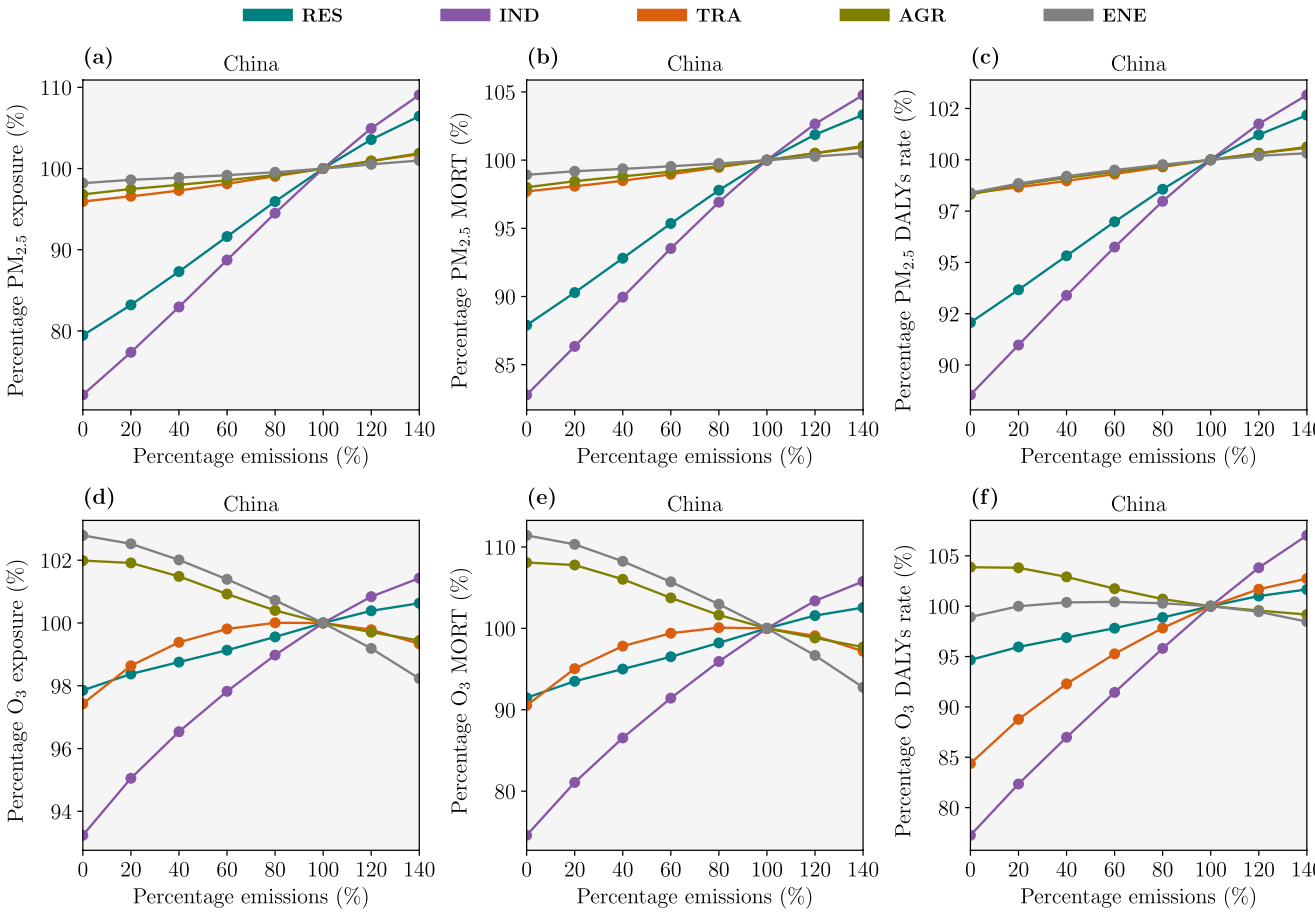

Figure 2. The impact of emission changes in China for 2015 on (a) fine particulate matter ($PM_{2.5}$, annual−mean) exposure, (b) annual premature mortalities (MORT) from $PM_{2.5}$ exposure, (c) annual rate of disability−adjusted life years per 100,000 people from $PM_{2.5}$ exposure, (d) ozone (O_3, maximum 6−monthly−mean daily−maximum 8−hour, 6mDM8h) exposure, (e) annual MORT from O_3 exposure, and (f) annual rate of DALYs per 100,000 people from O_3 exposure. Exposure, MORT, and rate of DALYs are shown relative to a simulation with baseline (100%) emissions. The emission sectors are residential, industry, land transport, agriculture, and power generation.

3.2. Impact of Changes in Individual Emission Sectors on $PM_{2.5}$ and O_3 Exposure and Disease Burden

$PM_{2.5}$ exposure decreases approximately linearly from emission reductions in a single sector (Figure 2a). Completely removing IND emissions decreases national $PM_{2.5}$ exposure by 28% and regional $PM_{2.5}$ exposure by 23%–31%. Under this scenario, the National Air Quality Target (35 μg m^{-3}) is achieved in all regions except North China (Figures S7−S13 in Supporting Information S1). Completely removing RES emissions decreases national $PM_{2.5}$ exposure by 21% and regional $PM_{2.5}$ exposure by 8%–28%. Removing IND or RES emissions results in similar reductions in regional $PM_{2.5}$ exposure, except in the GBA where reducing IND emissions provides larger reductions in $PM_{2.5}$ exposure. The impacts on $PM_{2.5}$ exposure from individual emission sector changes are then of decreasing size from TRA, AGR, and ENE emissions.

The sectoral contributions to $PM_{2.5}$ concentrations of 28% from industry, 21% from RES, and 4% from TRA emissions are similar to those from a multi−model study, which found contributions of 30% from industry, 26% from RES, and 7% from TRA emissions (Reddington et al., 2019). We find smaller contributions from AGR (3%) and ENE (2%) emissions compared to the multi−model contribution estimates of 16% from AGR and 14% from ENE emissions (Reddington et al., 2019). In our study, there are larger contributions from other sources (42%, compared to 7% in the multi−model estimates) (Reddington et al., 2019), such as from other anthropogenic emissions inside China, anthropogenic emissions outside China, and natural emissions.

The fractional reductions in $PM_{2.5}$ disease burden (Figure 2b) are smaller than the fractional reductions in $PM_{2.5}$ exposure (Figure 2a), due to the non−linear exposure−outcome association. Completely removing IND

emissions decreases the national number of premature mortalities from $PM_{2.5}$ exposure by 17%, avoiding 369,100 (95UI: 334,300–404,800) deaths, and decreases the rate of DALYs by 11%. Completely removing RES emissions decreases the national number of premature mortalities from $PM_{2.5}$ exposure by 12%, and decreases the rate of DALYs by 8%. Removing either IND or RES emissions can decrease the regional number of premature mortalities by 5%–20% and the rate of DALYs by 6%–19%.

O_3 exposure changes non−linearly with changes in emissions from a single sector (Figure 2d). This non−linear response is stronger for some sectors (e.g., TRA) and in some regions (e.g., North China, Figures S7−S13 in Supporting Information S1). Completely removing either IND or RES emissions consistently decreases O_3 exposure. Removing IND emissions decreases national O_3 exposure by 7% and regional O_3 exposure by 3%–10%. Removing RES emissions decreases national O_3 exposure by 2% and regional O_3 exposure by 0%–4%. However, individually removing ENE, TRA, or AGR emissions can increase O_3 exposure.

The fractional changes in the O_3 disease burden (Figure 2e) are larger than the fractional changes in O_3 exposure (Figure 2d), due to the high counterfactual exposure level of no excess risk. Completely removing IND emissions decreases the national number of premature mortalities from O_3 exposure by 25%, avoiding 11,700 (95UI: 8,100−16,400) deaths, and decreases the rate of DALYs by 23%. Completely removing RES emissions decreases the national number of premature mortalities from O_3 exposure by 9%, avoiding 3,900 (95UI: 2,700−5,500) deaths, and decreases the rate of DALYs by 5%. Although removing TRA emissions decreases the national number of premature mortalities from O_3 exposure by 9% avoiding 4,400 (95UI: 3,000–6,100) deaths, the number of premature mortalities in the GBA increases by 21%, an additional 300 (95UI: 200–400) deaths locally. Removing ENE emissions increases the national number of premature mortalities from O_3 exposure by 11%, whilst removing AGR emissions increases the national number of premature mortalities from O_3 exposure by 8%.

3.3. Impact of Changes in Multiple Emission Sectors on $PM_{2.5}$ and O_3 Exposure and Disease Burden

Removing RES and IND emissions together decreases national $PM_{2.5}$ exposure by 48% (Figure 3) and regional $PM_{2.5}$ exposure by 37%–53% (Figures S14−S20 in Supporting Information S1). These emission reductions attain the WHO Interim Target 2 (25 μg m^{-3}) in all regions except North and East China, and attain the WHO Interim Target 3 (15 μg m^{-3}) in the GBA. Removing IND emissions with either AGR, TRA, or ENE emissions attains the National Air Quality Target (35 μg m^{-3}) in China (Figure 3). Removing TRA, AGR, and ENE emissions, without reducing RES and IND emissions, decreases national $PM_{2.5}$ exposure by 9%.

Removing emissions from all five sectors together decreases national $PM_{2.5}$ exposure by 57% and regional $PM_{2.5}$ exposure by 52%–61%. These emission reductions attain the WHO Interim Target 2 (25 μg m^{-3}) in all regions except North China, but only attain the WHO Interim Target 3 (15 μg m^{-3}) in the GBA. The WHO Annual Guideline (5 μg m^{-3}) is not achieved in any region. Under this scenario, $PM_{2.5}$ concentrations result from other anthropogenic emissions inside China including shipping, aviation, and agricultural fires, from anthropogenic emissions outside China, and from natural emission sources such as vegetation fires, dust, sea spray, and secondary organic aerosols from biogenic volatile organic compounds (VOC). Previous studies have estimated the contributions to $PM_{2.5}$ concentrations in China from dust were 2%–10% and were higher in North West China (Shi et al., 2017; Yang et al., 2011). Open biomass burning was estimated to contribute 1%–8% of $PM_{2.5}$ concentrations across China (Reddington et al., 2019), with higher contributions of up to 29% in South Central China (Reddington et al., 2019; Shi et al., 2017). Biogenic SOA has been estimated to contribute 2%–8% of $PM_{2.5}$ concentrations (Hu et al., 2017; Shi et al., 2017). Long−range transport of $PM_{2.5}$ concentrations from anthropogenic emissions outside China was estimated to contribute up to 3% of $PM_{2.5}$ concentrations inside China (Liu et al., 2020). Anthropogenic emissions from shipping were estimated to contribute up to 3% (Chen et al., 2019; Dasadhikari et al., 2019; Reddington et al., 2019) and aviation 1% (Dasadhikari et al., 2019; Zhang et al., 2017). Emissions from sea salt have been estimated to contribute 1% of $PM_{2.5}$ concentrations (Shi et al., 2017).

Removing RES and IND emissions together decreases the national number of premature mortalities from $PM_{2.5}$ exposure by 32%, avoiding 691,800 (95UI: 625,100–760,400) deaths, and decreases the rate of DALYs by 21% (Figure 4 and S22 in Supporting Information S1). Removing TRA, AGR, and ENE emissions together decreases the national number of premature mortalities from $PM_{2.5}$ exposure by 5%. Removing emissions from all five sectors together decreases the national number of premature mortalities from $PM_{2.5}$ exposure by 40%, avoiding 858,800 (95UI: 774,900–945,400) deaths, and decreases the rate of DALYs by 27%.

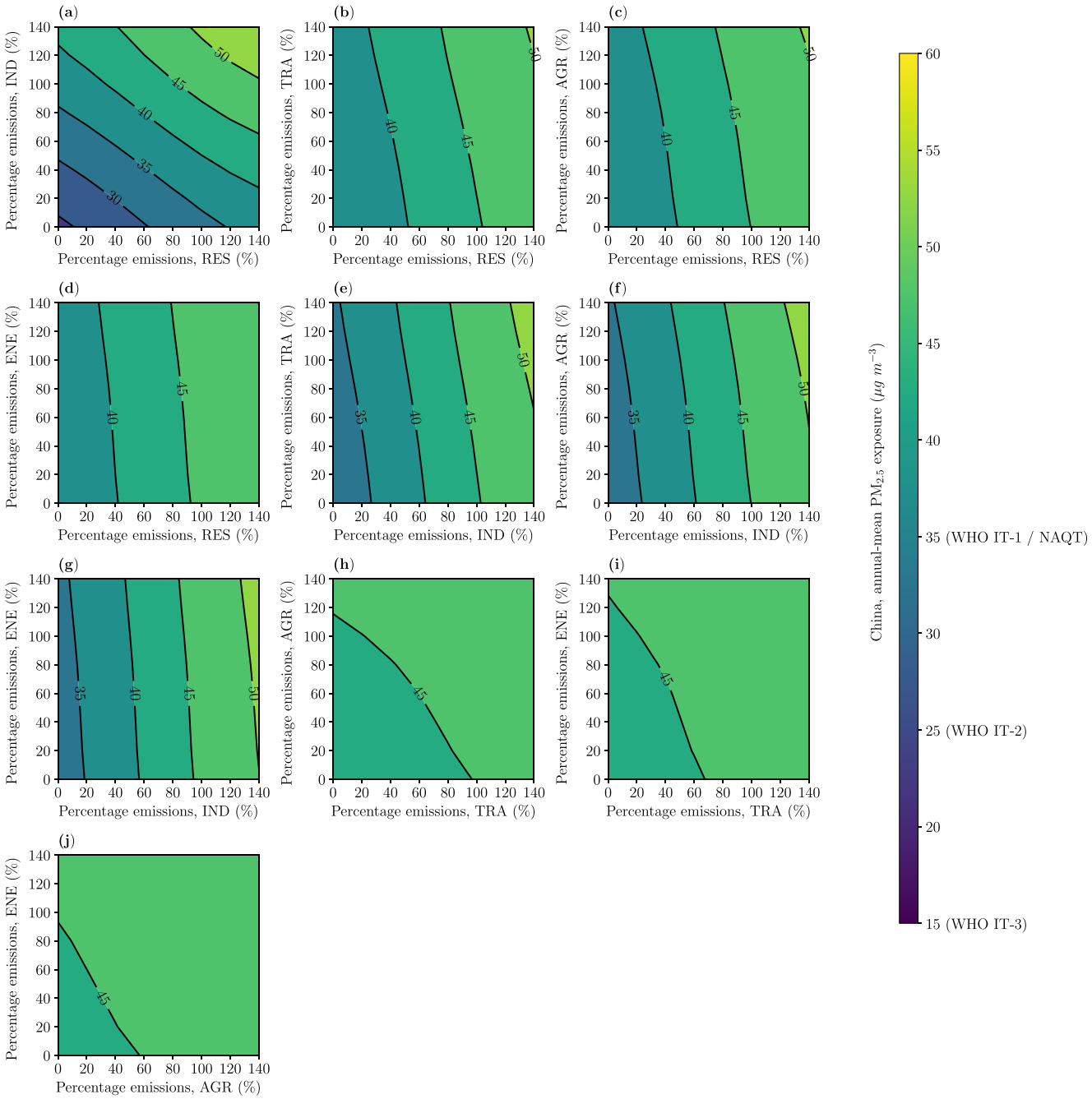

Figure 3. The impact of variations in two emission sectors on fine particulate matter ($PM_{2.5}$, annual−mean) exposure in China for 2015 from (a) residential (RES) and industry (IND), (b) RES and land transport (TRA), (c) RES and agriculture (AGR), (d) RES and power generation (ENE), (e) IND and TRA, (f) IND and AGR, (g) IND and ENE, (h) TRA and AGR, (i) TRA and ENE, and (j) AGR and ENE emissions. Air quality targets shown for the World Health Organization's (WHO) Interim Target 1 (IT−1, 35 μg m^{-3}), Interim Target 2 (IT−2, 25 μg m^{-3}), Interim Target 3 (IT−3, 15 μg m^{-3}), and China's National Air Quality Target (35 μg m^{-3}).

These findings highlight that substantial public health benefits can be achieved by emission reductions, and the majority of these benefits due to reductions in IND and RES emissions. However, even after removing emissions from these five sectors, approximately half of the disease burden associated with $PM_{2.5}$ exposure remains, due to other sources of air pollution.

The largest reductions in O_3 exposure occur when removing IND emissions with either TRA or RES emissions (Figure S22 in Supporting Information S1). Removing IND and TRA emissions decreases national O_3 exposure

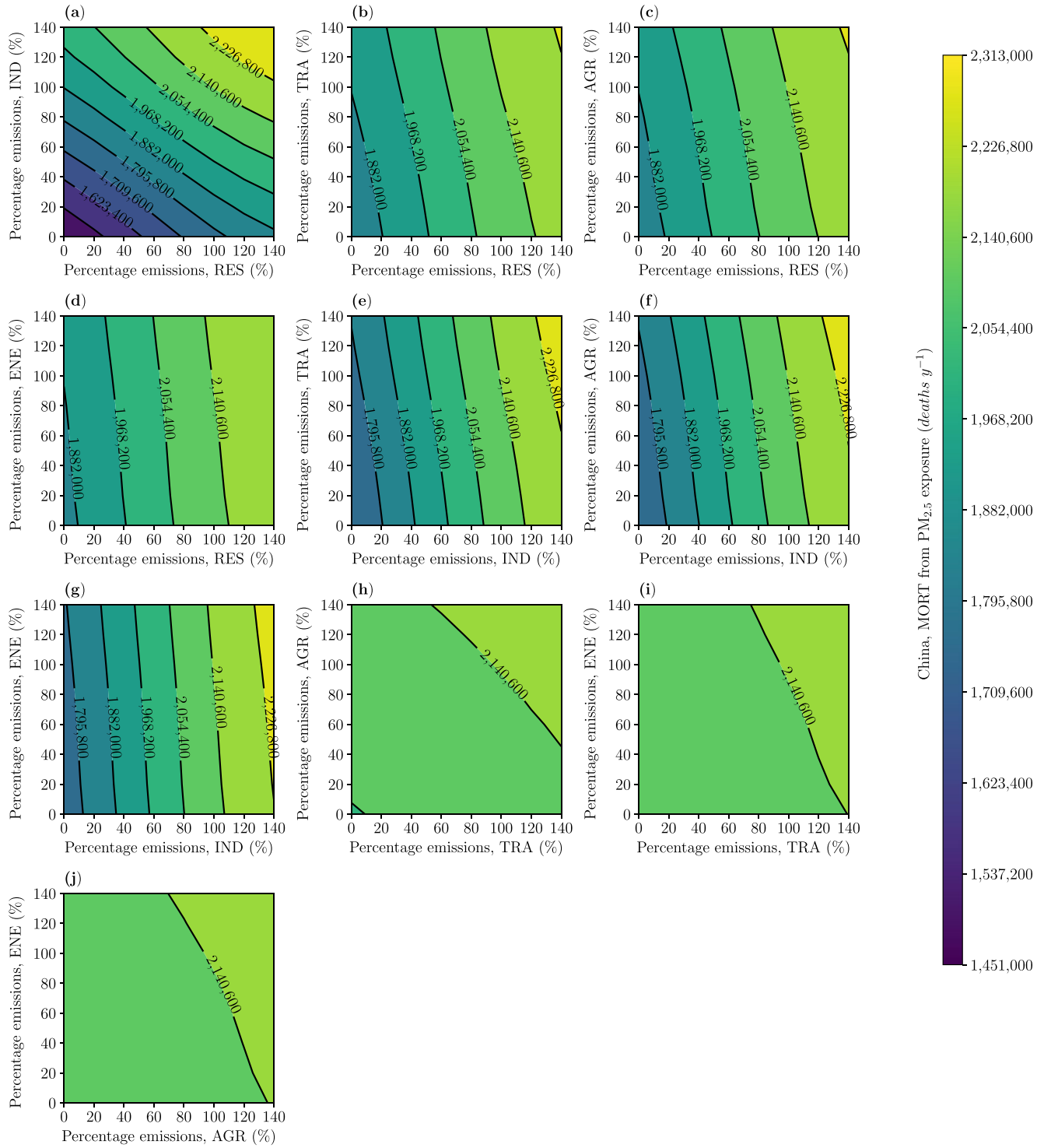

Figure 4. The impact of variations in two emission sectors on the disease burden (premature mortalities, MORT, per year) from fine particulate matter ($PM_{2.5}$, annual−mean) exposure for China from (a) residential (RES) and industry (IND), (b) RES and land transport (TRA), (c) RES and agriculture (AGR), (d) RES and power generation (ENE), (e) IND and TRA, (f) IND and AGR, (g) IND and ENE, (h) TRA and AGR, (i) TRA and ENE, and (j) AGR and ENE emissions.

by 14% and regional O_3 exposure by 8%–22% (Figures S23−S29 in Supporting Information S1), reducing the disease burden by 46% and the rate of DALYs by 41% (Figures S30 and S31 in Supporting Information S1). Removing IND and RES emissions decreases national O_3 exposure by 10% and regional O_3 exposure by 5%–16%

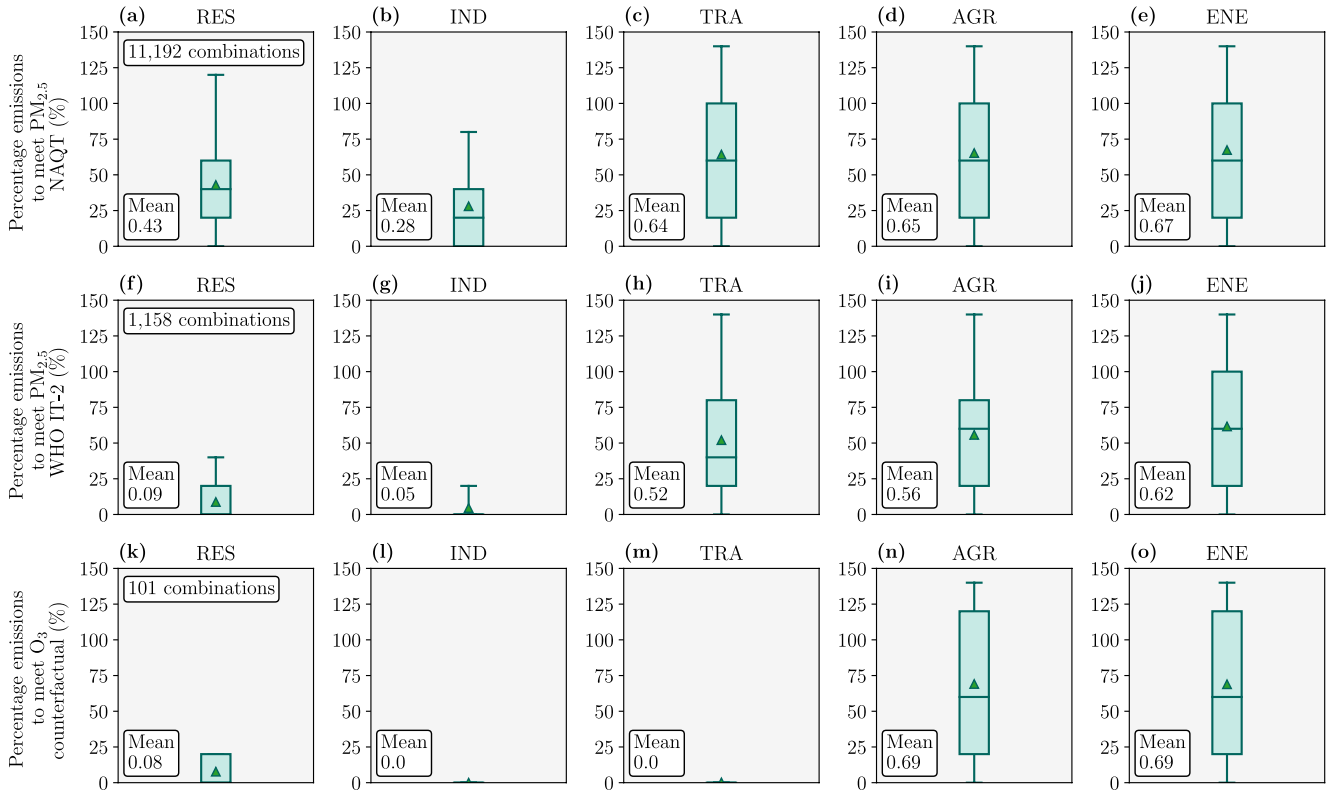

Figure 5. Emission configurations that meet air quality targets in China from the baseline in 2015. Targets are (a–e) the National Air Quality Target (35 μg m^{-3}) for ambient fine particulate matter ($PM_{2.5}$, annual−mean) concentrations (f–j) the World Health Organization Interim Target 2 (IT−2, 25 μg m^{-3}) for $PM_{2.5}$ concentrations, and (k–o) the counterfactual exposure level of no excess risk for ozone (O_3, 35.7 ppb) concentrations. Boxplots per sector from (a, f, and k) residential (RES), (b, g, and l) industrial (IND), (c, h, and m) land transport (TRA), (d, i, and n) agricultural (AGR), and (e, j, and o) power generation (ENE) emissions.

reducing the annual number of premature mortalities by 36%. However, some combinations of reductions in ENE, AGR, and TRA emissions can increase O_3 exposure and the associated disease burden in some regions.

These findings highlight the complex dependencies between the chemical production of O_3 and precursors of O_3, especially nitrogen oxides (NO_X) and VOC emissions. The ENE and TRA sectors have large NO_X emissions but smaller VOC emissions, while the IND sector has large emissions of both NO_X and VOC (Zheng et al., 2018). As some urban areas in East China are VOC−limited (Jin & Holloway, 2015; Wang et al., 2017), reducing NO_X emissions from the ENE and TRA sectors can increase O_3 exposure (Li et al., 2021).

The largest public health benefits come from reductions in $PM_{2.5}$ exposure, as the $PM_{2.5}$ disease burden is two orders of magnitude larger the O_3 disease burden. The largest public health benefits also come from reductions in IND and RES emissions, as these sectors dominate $PM_{2.5}$ exposure and are the only sectors which consistently decrease O_3 exposure.

3.4. Emission Configurations That Meet Air Quality Targets

There are 11,192 emission configurations that meet the National Air Quality Target (35 μg m^{-3}) nationally for $PM_{2.5}$ concentrations, requiring mean emission reductions of 72% in IND, 57% in RES, 36% in TRA, 35% in AGR, and 33% in ENE emissions (Figure 5). The WHO Interim Target 2 (25 μg m^{-3}) can be attained nationally for $PM_{2.5}$ concentrations via 1,158 emission configurations, requiring 95% reductions in IND and RES emissions. The GBA is the only region where the WHO Interim Target 3 (15 μg m^{-3}) can be attained, requiring stringent reductions in IND emissions (Figure S32 in Supporting Information S1). The WHO Air Quality Guideline (5 μg m^{-3}) cannot be attained in any region from reductions in these five emission sectors alone.

For O_3 concentrations, the National Air Quality Target (80 ppb) and the WHO guideline (50 ppb) are already attained at the baseline. There are 101 emission configurations that can reach the counterfactual exposure level of no excess risk (35.7 ppb) for O_3 concentrations, all of which require the near removal of IND, TRA, and RES emissions. Reaching this level would remove the health impacts associated with O_3 exposure. This attainment is possible in all regions except North and East China (Figure S32 in Supporting Information S1).

4. Conclusion

We developed emulators to predict annual mean $PM_{2.5}$ and O_3 concentrations and associated public health impacts in China. Our analysis provides a first estimate of how air pollution exposure and associated disease burden in China vary for different emission reductions across five major emission sectors (IND, RES, TRA, AGR, and ENE). The emulators predicted 99.9% of the variance in $PM_{2.5}$ and O_3 concentrations for a given input configuration of emissions. We found that $PM_{2.5}$ exposure was most sensitive to IND and RES emissions, confirming previous studies (Reddington et al., 2019). Complete removal of IND emissions would attain the National Air Quality Target (35 μg m^{-3}) in all regions except North China. The National Air Quality Target can be attained nationally with emission reductions of 72% in IND, 57% in RES, 36% in TRA, 35% in AGR, and 33% in ENE emissions. Removing RES and IND emissions completely, decreases national $PM_{2.5}$ exposure by 48% (to 24.0 μg m^{-3}). However, removing emissions from the five sectors in China does not enable the attainment of the WHO Annual Guideline (5 μg m^{-3}) due to the remaining emissions from shipping, aviation, and agricultural fires, emissions from outside China, and from natural emission sources including forest fires, dust, and biogenics.

Emulators have broad potential in air quality research. Future work could study other regions, countries, emission sources, and could further split emission inputs by species, sub−sectors, and time−of−day. Here, we chose to apply the same emission changes over all species within each sector due to computational constraints. For five inputs, 55 annual high–resolution WRFChem simulations were required for the training and testing data (Loeppky et al., 2009). If the emulators split the emissions by pollutant, then the computational burden would have increased by up to a factor of 10. Our work was conducted for one meteorological year (2015). Previous work found that the air quality impacts from meteorological changes were smaller than those from emission changes (Ding et al., 2019; Silver, Conibear, et al., 2020; Silver, He et al., 2020). However, future work should account for and explore variations in meteorology, including seasonal and inter−annual variations. To further understand how China can achieve the WHO Annual Guideline for $PM_{2.5}$ exposure (5 μg m^{-3}), future work is needed exploring the relative contributions of other anthropogenic emission sources, emissions outside China and natural emissions, which may increase under climate change (Carslaw et al., 2010). Our work highlights the challenges facing China as it attempts to further reduce $PM_{2.5}$ exposure and improve public health.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Code to setup and run WRFChem (using WRFotron version 2.0) is available through Conibear and Knote (2020). Emulator code and data is available through Conibear. (2021). The trained emulators per grid cell in China that support the findings of this study are available in Conibear et al. (2022b).

Acknowledgments
We gratefully acknowledge support from the AIA Group Limited, a European Research Council Consolidator Grant (771492), and the Natural Environment Research Council (NE/S006680/1). This work was undertaken on ARC4, part of the High−Performance Computing facilities at the University of Leeds, UK. This work used WRFotron version 2.0, a tool to automatise WRFChem runs with re–initialised meteorology (Conibear & Knote, 2020). We acknowledge the use of WRFChem preprocessor tools mozbc, fire_emiss, anthro_emiss, and bio_emiss provided by Atmospheric Chemistry Observations and Modeling (ACOM) of the National Center for Atmospheric Research (NCAR). We acknowledge the use of Model for Ozone and Related Chemical Tracers (MOZART) global model output, available at acom.ucar.edu/wrf−chem/mozart.shtml. We acknowledge the use of the emission pre–processor, available at github.com/douglowe/WRF_UoM_EMIT. We thank Qiang Zhang and Meng Li for providing MEIC data. We acknowledge the Python Software Foundation, Python Language Reference, available at python.org. We are particularly grateful to the Python libraries NumPy (Harris et al., 2020), Pandas (McKinney, 2010), Matplotlib (Hunter, 2007), SciPy (Virtanen et al., 2020), xarray (Hoyer & Hamman, 2017), Cartopy (Met Office, 2015), GeoPandas (Jordahl et al., 2020), Salem (Maussion, TimoRoth, Tbridel, Dusch, & Landmann, 2019), Jupyter (Kluyver et al., 2016), Scikit−learn (Pedregosa et al., 2011), TPOT (Olson et al., 2016), SALib (Herman & Usher, 2017), pyDOE, Seaborn (Waskom et al., 2020), Rasterio (Gillies, 2013), Affine, xESMF (Zhuang et al., 2020), Dask (Rocklin, 2015), and the Pangeo project (Abernathey et al., 2017). The boundaries shown on any maps in this work do not imply any judgement concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

References

Abernathey, R., Paul, K., Hamman, J., Rocklin, M., Lepore, C., Tippett, M., et al. (2017). *Pangeo NSF Earthcube proposal*. https://doi.org/10.6084/m9.figshare.5361094.v1

Aleksankina, K., Reis, S., Vieno, M., & Heal, M. R. (2019). Advanced methods for uncertainty assessment and global sensitivity analysis of an Eulerian atmospheric chemistry transport model. *Atmospheric Chemistry and Physics*, *19*(5), 2881–2898. https://doi.org/10.5194/acp-19-2881-2019

Beddows, A. V., Kitwiroon, N., Williams, M. L., & Beevers, S. D. (2017). Emulation and sensitivity analysis of the community multiscale Air quality model for a UK ozone pollution episode. *Environmental Science and Technology*, *51*(11), 6229–6236. https://doi.org/10.1021/acs.est.6b05873

Beucler, T., Pritchard, M., Gentine, P., & Rasp, S. (2020). Towards physically-consistent, data-driven models of convection. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 3987–3990. Section 3. https://doi.org/10.1109/IGARSS39084.2020.9324569

Beusch, L., Gudmundsson, L., & Seneviratne, S. I. (2020). Emulating Earth system model temperatures with MESMER: From global mean temperature trajectories to grid-point-level realizations on land. *Earth System Dynamics*, *11*(1), 139–159. https://doi.org/10.5194/esd-11-139-2020

Brasseur, G. P., & Jacob, D. J. (2017). *Modeling of atmospheric chemistry (first)*. Cambridge University Press. https://doi.org/10.1017/9781316544754

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*(12), 6289–6298. https://doi.org/10.1029/2018GL078510

Buonocore, J. J., Dong, X., Spengler, J. D., Fu, J. S., & Levy, J. I. (2014). Using the Community Multiscale Air Quality (CMAQ) model to estimate public health impacts of $PM_{2.5}$ from individual power plants. *Environment International*, *68*, 200–208. https://doi.org/10.1016/j.envint.2014.03.031

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C. A., et al. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences*, *115*(38), 9592–9597. https://doi.org/10.1073/pnas.1803222115

Carnevale, C., Finzi, G., Pisoni, E., & Volta, M. (2009). Neuro-fuzzy and neural network systems for air quality control. *Atmospheric Environment*, *43*(31), 4811–4821. https://doi.org/10.1016/j.atmosenv.2008.07.064

Carslaw, K. S., Boucher, O., Spracklen, D. V., Mann, G. W., Rae, J. G. L., Woodward, S., & Kulmala, M. (2010). A review of natural aerosol interactions and feedbacks within the Earth system. *Atmospheric Chemistry and Physics*, *10*(4), 1701–1737. https://doi.org/10.5194/acp-10-1701-2010

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., et al. (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature*, *503*(7474), 67–71. https://doi.org/10.1038/nature12674

Center for International Earth Science Information Network, & NASA Socioeconomic Data and Applications Center. (2016). *Gridded population of the World, version 4 (GPWv4): Population count*. Columbia University. https://doi.org/10.7927/H4NP22DQ

Chang, W., Haran, M., Applegate, P., & Pollard, D. (2016). Calibrating an ice sheet model using high-dimensional binary spatial data. *Journal of the American Statistical Association*, *111*(513), 57–72. https://doi.org/10.1080/01621459.2015.1108199

Chantry, M., Christensen, H., Dueben, P., & Palmer, T. (2021). Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *379*(2194), 20200083. https://doi.org/10.1098/rsta.2020.0083

Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I., & Palmer, T. (2021). Machine learning emulation of gravity wave drag in numerical weather forecasting. *Journal of Advances in Modeling Earth Systems*, *13*(7), 1–20. https://doi.org/10.1029/2021MS002477

Chen, C., Saikawa, E., Comer, B., Mao, X., & Rutherford, D. (2019). Ship emission impacts on air quality and human health in the Pearl river delta (PRD) region, China, in 2015, with projections to 2030. *GeoHealth*, *3*(9), 284–306. https://doi.org/10.1029/2019GH000183

Chen, Y., Wild, O., Ryan, E., Sahu, S. K., Lowe, D., Archer-Nicholls, S., et al. (2020). Mitigation of $PM_{2.5}$ and ozone pollution in Delhi: A sensitivity study during the pre-monsoon period. *Atmospheric Chemistry and Physics*, *20*(1), 499–514. https://doi.org/10.5194/acp-20-499-2020

Cheng, J., Su, J., Cui, T., Li, X., Dong, X., Sun, F., et al. (2019). Dominant role of emission reduction in $PM_{2.5}$ air quality improvement in Beijing during 2013-2017: A model-based decomposition analysis. *Atmospheric Chemistry and Physics*, *19*(9), 6125–6146. https://doi.org/10.5194/acp-19-6125-2019

Chevallier, F., Morcrette, J. J., Chéruy, F., & Scott, N. A. (2000). Use of a neural-network-based long-wave radiative-transfer scheme in the ECMWF atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, *126*(563), 761–776. https://doi.org/10.1002/qj.49712656318

Conibear, L. (2021). lukeconibear/emulator_annual: Long-term emulator of WRFChem (Version v1.0.0). [Software]. Zenodo. https://doi.org/10.5281/zenodo.5549368

Conibear, L., & Knote, C. (2020). wrfchem-leeds/WRFotron: WRFotron 2.0. [Software]. Zenodo. https://doi.org/10.5281/zenodo.3624087

Conibear, L., Reddington, C. L., Silver, B. J., Chen, Y., Arnold, S. R., & Spracklen, D. V. (2022a). Emission Sector Impacts on Air Quality and Public Health in China from 2010−2020. *GeoHealth*. https://doi.org/10.1029/2021GH000567

Conibear, L., Reddington, C. L., Silver, B. J., Chen, Y., Arnold, S. R., & Spracklen, D. V. (2022b). *Supplementary data: Sensitivity of air pollution exposure and disease burden to emission changes in China using machine learning emulation.* University of Leeds. [Dataset]. https://doi.org/10.5518/1055

Conibear, L., Reddington, C. L., Silver, B. J., Chen, Y., Knote, C., Arnold, S. R., & Spracklen, D. V. (2021). Statistical emulation of winter ambient fine particulate matter concentrations from emission changes in China. *GeoHealth*, *5*. https://doi.org/10.1029/2021GH000391

Conibear, L., Reddington, C. L., Silver, B. J., Knote, C., Arnold, S. R., & Spracklen, D. V. (2021). Regional policies targeting residential solid fuel and agricultural emissions can improve air quality and public health in the greater Bay area and across China. *GeoHealth*, *5*(4). https://doi.org/10.1029/2020GH000341

Dagon, K., Sanderson, B. M., Fisher, R. A., & Lawrence, D. M. (2020). A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5. *Advances in Statistical Climatology, Meteorology and Oceanography*, *6*(2), 223–244. https://doi.org/10.5194/ascmo-6-223-2020

Dasadhikari, K., Eastham, S. D., Allroggen, F., Speth, R. L., & Barrett, S. R. H. (2019). Evolution of sectoral emissions and contributions to mortality from particulate matter exposure in the Asia-Pacific region between 2010 and 2015. *Atmospheric Environment*, *216*, 116916. https://doi.org/10.1016/j.atmosenv.2019.116916

Deutsch, D. (2012). *The beginning of infinity: Explanations that transform the World*. Penguin.

Ding, D., Xing, J., Wang, S., Liu, K., & Hao, J. (2019). Estimated contributions of emissions controls, meteorological factors, population growth, and changes in baseline mortality to reductions in ambient $PM_{2.5}$ and $PM_{2.5}$-related mortality in China, 2013–2017. *Environmental Health Perspectives*, *127*(6), 1–12. https://doi.org/10.1289/EHP4157

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

Fleming, Z. L., Doherty, R. M., Von Schneidemesser, E., Malley, C. S., Cooper, O. R., Pinto, J. P., et al. (2018). Tropospheric ozone assessment report: Present-day ozone distribution and trends relevant to human health. *Elementa: Science of the Anthropocene*, *6*(12), 41. https://doi.org/10.1525/elementa.273

Foley, K. M., Napelenok, S. L., Jang, C., Phillips, S., Hubbell, B. J., & Fulcher, C. M. (2014). Two reduced form air quality modeling techniques for rapidly calculating pollutant mitigation potential across many sources, locations and precursor emission types. *Atmospheric Environment*, *98*, 283–289. https://doi.org/10.1016/j.atmosenv.2014.08.046

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742–5751. https://doi.org/10.1029/2018GL078202

Gettelman, A., Gagne, D. J., Chen, C. C., Christensen, M. W., Lebo, Z. J., Morrison, H., & Gantos, G. (2021). Machine learning the warm rain process. *Journal of Advances in Modeling Earth Systems*, *13*(2). https://doi.org/10.1029/2020MS002268

Gillies, S. (2013). *Rasterio: Geospatial raster I/O for Python programmers*. Retrieved from https://github.com/mapbox/rasterio

Global Burden of Disease Study 2017 (2018). *Global burden of disease study 2017 (GBD 2017) population estimates 1950-2017*. Retrieved from http://ghdx.healthdata.org/record/ihme-data/gbd-2017-population-estimates-1950-2017

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, *39*(37), 6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027

Han, Y., Zhang, G. J., Huang, X., & Wang, Y. (2020). A moist physics parameterization based on deep learning. *Journal of Advances in Modeling Earth Systems*, *12*(9). https://doi.org/10.1029/2020MS002076

Harder, P., Watson-Parris, D., Strassel, D., Gauger, N., Stier, P., & Keuper, J. (2021). *Emulating aerosol microphysics with a machine learning*. Retrieved from http://arxiv.org/abs/2109.10593

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hatfield, S., Chantry, M., Dueben, P., Lopez, P., Geer, A., & Palmer, T. (2021). Building tangent-linear and adjoint models for data assimilation with neural networks. *Journal of Advances in Modeling Earth Systems*, *13*(9), 1–16. https://doi.org/10.1029/2021MS002521

Henze, D. K., Hakami, A., & Seinfeld, J. H. (2007). Development of the adjoint of GEOS-Chem. *Atmospheric Chemistry and Physics*, *7*(9), 2413–2433. https://doi.org/10.5194/acp-7-2413-2007

Heo, J., Adams, P. J., & Gao, H. O. (2016a). Public health costs of primary $PM_{2.5}$ and inorganic $PM_{2.5}$ precursor emissions in the United States. *Environmental Science and Technology*, *50*(11), 6061–6070. https://doi.org/10.1021/acs.est.5b06125

Heo, J., Adams, P. J., & Gao, H. O. (2016b). Reduced-form modeling of public health impacts of inorganic $PM_{2.5}$ and precursor emissions. *Atmospheric Environment*, *137*, 80–89. https://doi.org/10.1016/j.atmosenv.2016.04.026

Herman, J., & Usher, W. (2017). SALib: An open-source Python library for sensitivity analysis. *Journal of Open Source Software*, *2*(9), 97. https://doi.org/10.21105/joss.00097

Hijmans, R., University of California, Berkeley Museum of Vertebrate Zoology, Kapoor, J., Wieczorek, J., & International Rice Research Institute (2020). *Global administrative areas (GADM): Boundaries without limits. Version 3.6*. Retrieved from http://gadm.org/

Hodzic, A., & Jimenez, J. L. (2011). Modeling anthropogenically controlled secondary organic aerosols in a megacity: A simplified framework for global and climate models. *Geoscientific Model Development*, *4*(4), 901–917. https://doi.org/10.5194/gmd-4-901-2011

Hodzic, A., & Knote, C. (2014). *WRF-chem 3.6.1: MOZART gas-phase chemistry with MOSAIC aerosols* (Vol. 7). Atmospheric Chemistry Division (ACD), National Center for Atmospheric Research (NCAR).

Holden, P. B., Edwards, N. R., Rangel, T. F., Pereira, E. B., Tran, G. T., & Wilkinson, R. D. (2019). PALEO-PGEM v1.0: A statistical emulator of pliocene-pleistocene climate. *Geoscientific Model Development*, *12*(12), 5137–5155. https://doi.org/10.5194/gmd-12-5137-2019

Hoyer, S., & Hamman, J. J. (2017). xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*, *5*, 1–6. https://doi.org/10.5334/jors.148

Hu, J., Wang, P., Ying, Q., Zhang, H., Chen, J., Ge, X., et al. (2017). Modeling biogenic and anthropogenic secondary organic aerosol in China. *Atmospheric Chemistry and Physics*, *17*(1), 77–92. https://doi.org/10.5194/acp-17-77-2017

Hughes, M., Kodros, J. K., Pierce, J. R., West, M., & Riemer, N. (2018). Machine learning to predict the global distribution of aerosol mixing state metrics. *Atmosphere*, *9*(15), 1–18. https://doi.org/10.3390/atmos9010015

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Institute for Health Metrics and Evaluation. (2020). *GBD compare data visualization*. Retrieved from healthdata.org/gbd-compare

Ivatt, P. D., & Evans, M. J. (2020). Improving the prediction of an atmospheric chemistry transport model using gradient-boosted regression trees. *Atmospheric Chemistry and Physics*, *20*(13), 8063–8082. https://doi.org/10.5194/acp-20-8063-2020

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., et al. (2015). HTAP-v2.2: A mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution. *Atmospheric Chemistry and Physics*, *15*(19), 11411–11432. https://doi.org/10.5194/acp-15-11411-2015

Jin, L., Wang, B., Shi, G., Seyler, B. C., Qiao, X., Deng, X., et al. (2020). Impact of China's recent amendments to air quality monitoring protocol on reported trends. *Atmosphere*, *11*(1199), 1–8. https://doi.org/10.3390/atmos11111199

Jin, X., & Holloway, T. (2015). Spatial and temporal variability of ozone sensitivity over China observed from the Ozone Monitoring Instrument. *Journal of Geophysical Research: Atmospheres*, *120*(14), 7229–7246. https://doi.org/10.1002/2015JD023250

Johnson, J. S., Cui, Z., Lee, L. A., Gosling, J. P., Blyth, A. M., & Carslaw, K. S. (2015). Evaluating uncertainty in convective cloud microphysics using statistical emulation. *Journal of Advances in Modeling Earth Systems*, *7*(1), 162–187. https://doi.org/10.1002/2014MS000383

Jordahl, K., Bossche, J., Vanden Wasserman, J., McBride, J., Fleischmann, M., & Gerard, J. (2020). geopandas/geopandas: V0.7.0 (Version v0.7.0). *Zenodo*. https://doi.org/10.5281/zenodo.3669853

Kasim, M. F., Watson-Parris, D., Deaconu, L., Oliver, S., Hatfield, P., Froula, D. H., et al. (2022). Building high accuracy emulators for scientific simulations with deep neural architecture search. *Machine Learning: Science and Technology*, *3*(1), 015013. https://doi.org/10.1088/2632-2153/ac3ffa

Keller, C. A., & Evans, M. J. (2019). Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10. *Geoscientific Model Development*, *12*(3), 1209–1225. https://doi.org/10.5194/gmd-12-1209-2019

Kelp, M. M., Jacob, D. J., Kutz, J. N., Marshall, J. D., & Tessum, C. W. (2020). Toward stable, general machine-learned models of the atmospheric chemical system. *Journal of Geophysical Research: Atmospheres*, *125*(23), e2020JD032759. https://doi.org/10.1029/2020JD032759

Kelp, M. M., Jacob, D. J., Lin, H., & Sulprizio, M. P. (2022). An online-learned neural network chemical solver for stable long-term global simulations of atmospheric chemistry. *Earth*, 1–25. https://doi.org/10.31223/X52K7J

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., et al. (2016). Jupyter notebooks – A publishing format for reproducible computational workflows. https://doi.org/10.3233/978-1-61499-649-1-87

Knote, C., Hodzic, A., & Jimenez, J. L. (2015). The effect of dry and wet deposition of condensable vapors on secondary organic aerosols concentrations over the continental US. *Atmospheric Chemistry and Physics*, *15*(1), 1–18. https://doi.org/10.5194/acp-15-1-2015

Knote, C., Hodzic, A., Jimenez, J. L., Volkamer, R., Orlando, J. J., Baidar, S., et al. (2014). Simulation of semi-explicit mechanisms of SOA formation from glyoxal in aerosol in a 3-D model. *Atmospheric Chemistry and Physics*, *14*(12), 6213–6239. https://doi.org/10.5194/acp-14-6213-2014

Krasnopolsky, V. (2020). *Using machine learning for model physics: An overview, (January)*. Retrieved from http://arxiv.org/abs/2002.00416

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Chalikov, D. V. (2005). 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*(5), 1370–1383. https://doi.org/10.1175/MWR2923.1

Lee, L. A., Carslaw, K. S., Pringle, K. J., & Mann, G. W. (2012). Mapping the uncertainty in global CCN using emulation. *Atmospheric Chemistry and Physics*, *12*(20), 9739–9751. https://doi.org/10.5194/acp-12-9739-2012

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., & Spracklen, D. V. (2011). Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmospheric Chemistry and Physics*, *11*(23), 12253–12273. https://doi.org/10.5194/acp-11-12253-2011

Lee, L. A., Reddington, C. L., & Carslaw, K. S. (2016). On the relationship between aerosol model uncertainty and radiative forcing uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5820–5827. https://doi.org/10.1073/pnas.1507050113

Li, K., Jacob, D. J., Liao, H., Qiu, Y., Shen, L., Zhai, S., et al. (2021). Ozone pollution in the North China Plain spreading into the late-winter haze season. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10), 1–7. https://doi.org/10.1073/pnas.2015797118

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, K., Jacob, D. J., Liao, H., Zhu, J., Shah, V., Shen, L., et al. (2020). A two-pollutant strategy for improving ozone and particulate air quality in China. *Nature Geoscience*, *12*(11), 906–910. https://doi.org/10.1038/s41561-019-0464-x

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

Li, M., Zhang, Q., Kurokawa, J., Woo, J.-H., He, K., Lu, Z., et al. (2017). Mix: A mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP. *Atmospheric Chemistry and Physics*, *17*(2), 935–963. https://doi.org/10.5194/acp-17-935-2017

Liu, S., Xing, J., Wang, S., Ding, D., Chen, L., & Hao, J. (2020). Revealing the impacts of transboundary pollution on $PM_{2.5}$-related deaths in China. *Environment International*, *134*, 105323. https://doi.org/10.1016/j.envint.2019.105323

Loeppky, J. L., Sacks, J., & Welch, W. J. (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, *51*(4), 366–376. https://doi.org/10.1198/TECH.2009.08040

Maussion, F., TimoRoth, T., Dusch, M., & Landmann, J. (2019). fmaussion/salem: V0.2.4 (Version v0.2.4). *Zenodo*. https://doi.org/10.5281/zenodo.2605265

McCoy, I. L., McCoy, D. T., Wood, R., Regayre, L., Watson-Parris, D., Grosvenor, D. P., et al. (2020). The hemispheric contrast in cloud microphysical properties constrains aerosol forcing. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(32), 18998–19006. https://doi.org/10.1073/pnas.1922502117

McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56.

Meic Research Group, & Tsinghua University (2019). *Multi-resolution emission inventory for China (MEIC) version 1.3*. Retrieved from http://www.meicmodel.org/

Met Office. (2015). Cartopy: A cartographic python library with a matplotlib interface. Retrieved from http://scitools.org.uk/cartopy

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), 1135–1159. https://doi.org/10.1016/S0140-6736(20)30752-2

Nethery, R. C., & Dominici, F. (2019). Estimating pollution-attributable mortality at the regional and global scales: Challenges in uncertainty estimation and causal inference. *European Heart Journal*, *0*(20), 1–3. https://doi.org/10.1093/eurheartj/ehz200

Nicely, J., Duncan, B., Hanisco, T., Wolfe, G., Salawitch, R., Deushi, M., et al. (2020). A machine learning examination of hydroxyl radical differences among model simulations for CCMI-1. *Atmospheric Chemistry and Physics*, *20*(3), 1341–1361. https://doi.org/10.5194/acp-2019-772

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*, 2548–2563. https://doi.org/10.1029/2018MS001351

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, *91*(10–11), 1290–1300. https://doi.org/10.1016/j.ress.2005.11.025

Olson, R. S., Bartley, N., Urbanowicz, R. J., & Moore, J. H. (2016). Evaluation of a tree-based pipeline optimization tool for automating data science. *GECCO 2016 - Proceedings of the 2016 Genetic and Evolutionary Computation Conference*, 485–492. https://doi.org/10.1145/2908812.2908918

Ott, J., Pritchard, M., Best, N., Linstead, E., Curcic, M., & Baldi, P. (2020). A fortran-keras deep learning bridge for scientific computing. *Scientific Programming*, 1–16. https://doi.org/10.1155/2020/8888811

Palmer, T. (2015). Build imprecise supercomputers. *Nature*, *526*(7571), 2–3. https://doi.org/10.1038/526032a

Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, *62*(3), 54–60. https://doi.org/10.1145/3241036

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python fabian. *Journal OfMachine Learning Research*, *12*, 2825–2830.

Qi, J., Zheng, B., Li, M., Yu, F., Chen, C., Liu, F., et al. (2017). A high-resolution air pollutants emission inventory in 2013 for the Beijing-Tianjin-Hebei region, China. *Atmospheric Environment*, *170*, 156–168. https://doi.org/10.1016/j.atmosenv.2017.09.039

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. The MIT Press. Massachusetts Institute of Technology.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Reddington, C., Conibear, L., Silver, B., Silver, B. J., Knote, C., Li, Y. J., et al. (2019). Exploring the impacts of anthropogenic emission sectors on $PM_{2.5}$ and human health in South and East Asia. *Atmospheric Chemistry and Physics*, *19*(18), 11887–11910. https://doi.org/10.5194/acp-19-11887-2019

Rocklin, M. (2015). Dask: Parallel computation with blocked algorithms and Task scheduling. *Proceedings of the 14th Python in science conference, (SCIPY)*, 126–132. https://doi.org/10.25080/majora-7b98e3ed-013

Ryan, E., & Wild, O. (2021). Calibrating a global atmospheric chemistry transport model using Gaussian process emulation and ground-level concentrations of ozone and carbon monoxide. *Geoscientific Model Development Discussions*, *14*(9), 5373–5391. https://doi.org/10.5194/gmd-2021-39

Ryan, E., Wild, O., Voulgarakis, A., & Lee, L. (2018). Fast sensitivity analysis methods for computationally expensive models with multi-dimensional output. *Geoscientific Model Development*, *11*(8), 3131–3146. https://doi.org/10.5194/gmd-11-3131-2018

Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine learning approaches for outdoor air quality modelling: A systematic review. *Applied Sciences*, *8*(12), 2570. https://doi.org/10.3390/app8122570

Salter, J. M., Williamson, D. B., Gregoire, L. J., & Edwards, T. L. (2018). *Quantifying spatio-temporal boundary condition uncertainty for the North American deglaciation*. ArXiv.

Scher, S. (2018). Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophysical Research Letters*, *45*(22), 12616–12622. https://doi.org/10.1029/2018GL080704

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics. From air pollution to climate change*. John Wiley & Sons, Inc. Third.

Shi, Z., Li, J., Huang, L., Wang, P., Wu, L., Ying, Q., et al. (2017). Source apportionment of fine particulate matter in China in 2013 using a source-oriented chemical transport model. *The Science of the Total Environment*, *601–602*, 1476–1487. https://doi.org/10.1016/j.scitotenv.2017.06.019

Silva, S. J., Ma, P. L., Hardin, J. C., & Rothenberg, D. (2021). Physically regularized machine learning emulators of aerosol activation. *Geoscientific Model Development*, *14*(5), 3067–3077. https://doi.org/10.5194/gmd-14-3067-2021

Silver, B., Conibear, L., Reddington, C. L., Knote, C., Arnold, S. R., & Spracklen, D. V. (2020). Pollutant emission reductions deliver decreased $PM_{2.5}$-caused mortality across China during 2015–2017. *Atmospheric Chemistry and Physics*, *20*, 11683–11695. https://doi.org/10.5194/acp-20-11683-2020

Silver, B., He, X., Arnold, S. R., & Spracklen, D. V. (2020). The impact of COVID-19 control measures on air quality in China. *Environmental Research Letters*, *15*(8), 084021. https://doi.org/10.1088/1748-9326/aba3a2

Silver, B., Reddington, C. L., Arnold, S. R., & Spracklen, D. V. (2018). Substantial changes in air pollution across China during 2015 to 2017. *Environmental Research Letters*, *13*(11), 114012. https://doi.org/10.1088/1748-9326/aae718

Skamarock, W. C., Klemp, J. B., Dudhi, J., Gill, D. O., Barker, D. M., Duda, M. G., et al. (2008). A description of the advanced research WRF version 3. *NCAR Tech*, *468+STR*, 113.

Stanaway, J. D., Afshin, A., Gakidou, E., Lim, S. S., Abate, D., Abate, K. H., et al. (2018). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: A systematic analysis for the global burden of disease stu. *The Lancet*, *392*(10159), 1923–1994. https://doi.org/10.1016/S0140-6736(18)32225-6

Tessum, C. W., Hill, J. D., & Marshall, J. D. (2017). InMAP: A model for air pollution interventions. *PLoS One*, *12*(4), 1–26. https://doi.org/10.1371/journal.pone.0176131

Tran, G. T., Oliver, K. I. C., Sóbester, A., Toal, D. J. J., Holden, P. B., Marsh, R., et al. (2016). Building a traceable climate model hierarchy with multi-level emulators. *Advances in Statistical Climatology, Meteorology and Oceanography*, *2*(1), 17–37. https://doi.org/10.5194/ascmo-2-17-2016

Turner, M. C., Jerrett, M., Pope, C. A., III, Krewski, D., Gapstur, S. M., Diver, R. W., et al. (2016). Long-term ozone exposure and mortality in a large prospective study. *American Journal of Respiratory and Critical Care Medicine*, *193*(10), 1134–1142. https://doi.org/10.1164/rccm.201508-1633OC

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*.

Wang, T., Xue, L., Brimblecombe, P., Lam, Y. F., Li, L., & Zhang, L. (2017). Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. *The Science of the Total Environment*, *575*, 1582–1596. https://doi.org/10.1016/j.scitotenv.2016.10.081

Waskom, M., Botvinnik, O., Ostblom, J., Gelbart, M., Lukauskas, S., & Hobson, P. (2020). mwaskom/seaborn: V0.10.1 (April 2020) (Version v0.10.1). *Zenodo*. https://doi.org/10.5281/zenodo.3767070

Watson-Parris, D. (2021). Machine learning for weather and climate are worlds apart. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *379*(2194), 20200098. https://doi.org/10.1098/rsta.2020.0098

Watson-Parris, D., Bellouin, N., Deaconu, L. T., Schutgens, N. A. J., Yoshioka, M., Regayre, L. A., et al. (2020). Constraining uncertainty in aerosol direct forcing. *Geophysical Research Letters*, *47*(9), 1–7. https://doi.org/10.1029/2020GL087141

Watson-Parris, D., Williams, A., Deaconu, L., & Stier, P. (2021). Model Calibration Using ESEm v1.0.0 – an Open. Scalable Earth System Emulator. *Geoscientific Model Development*, *14*(12), 7659–7672. https://doi.org/10.5194/gmd-14-7659-2021

Weyn, J. A., Durran, D. R., & Caruana, R. (2019). Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2680–2693. https://doi.org/10.1029/2019MS001705

Wild, O., Voulgarakis, A., O'Connor, F., Lamarque, J.-F., Ryan, E. M., & Lee, L. (2020). Global sensitivity analysis of chemistry–climate model budgets of tropospheric ozone and OH: Exploring model diversity. *Atmospheric Chemistry and Physics Discussions*, *20*(7), 4047–4058. https://doi.org/10.5194/acp-20-4047-2020

World Health Organization (2021). *WHO global air quality guidelines - particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. Retrieved from https://apps.who.int/iris/handle/10665/345329

Yang, F., Tan, J., Zhao, Q., Du, Z., He, K., Ma, Y., et al. (2011). Characteristics of $PM_{2.5}$ speciation in representative megacities and across China. *Atmospheric Chemistry and Physics*, *11*(11), 5207–5219. https://doi.org/10.5194/acp-11-5207-2011

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959. https://doi.org/10.1093/biomet/87.4.954

Yin, P., Brauer, M., Cohen, A. J., Wang, H., Li, J., Burnett, R. T., et al. (2020). The effect of air pollution on deaths, disease burden, and life expectancy across China and its provinces, 1990–2017: An analysis for the global burden of disease study 2017. *The Lancet Planetary Health*, *5196*(20), 1–13. https://doi.org/10.1016/S2542-5196(20)30161-3

Yu, S., Eder, B., Dennis, R., Chu, S.-H., & Schwartz, S. E. (2006). New unbiased symmetric metrics for evaluation of air quality models. *Atmospheric Science Letters*, *7*(1), 26–34. https://doi.org/10.1002/asl.125

Yuval, J., & O'Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-17142-3

Zaveri, R. A., Easter, R. C., Fast, J. D., & Peters, L. K. (2008). Model for simulating aerosol interactions and chemistry (MOSAIC). *Journal of Geophysical Research*, *113*(D13204), 1–29. https://doi.org/10.1029/2007JD008782

Zhai, S., Jacob, D. J., Wang, X., Shen, L., Li, K., Zhang, Y., et al. (2019). Fine particulate matter ($PM_{2.5}$) trends in China, 2013-2018: Contributions from meteorology. *Atmospheric Chemistry and Physics Discussions*, 1–19. https://doi.org/10.5194/acp-2019-279

Zhang, Q., Jiang, X., Tong, D., Davis, S. J., Zhao, H., Geng, G., et al. (2017). Transboundary health impacts of transported global air pollution and international trade. *Nature*, *543*(7647), 705–709. https://doi.org/10.1038/nature21712

Zhao, B., Zheng, H., Wang, S., Smith, K. R., Lu, X., Aunan, K., et al. (2018). Change in household fuels dominates the decrease in $PM_{2.5}$ exposure and premature mortality in China in 2005–2015. *Proceedings of the National Academy of Sciences*, *115*(49), 12401–12406. https://doi.org/10.1073/pnas.1812955115

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., Tong, D., Chen, C., Hong, C., Li, M., et al. (2017). Resolution dependence of uncertainties in gridded emission inventories: A case study in Hebei, China. *Atmospheric Chemistry and Physics*, *17*(2), 921–933. https://doi.org/10.5194/acp-17-921-2017

Zhuang, J., Dussin, R., Jüling, A., & Rasp, S. (2020). JiaweiZhuang/xESMF: V0.3.0 adding ESMF.LocStream capabilities (version v0.3.0). *Zenodo*. https://doi.org/10.5281/zenodo.3700105