



Statistical and machine learning methods for evaluating trends in air quality under changing meteorological conditions

Minghao Qiu^{1,a}, Corwin Zigler², and Noelle E. Selin^{1,3}

¹Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

²Department of Statistics and Data Science, University of Texas at Austin, Texas, USA

³Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

^acurrent address: Earth System Science Department, Stanford University, Stanford, California, USA

Correspondence: Minghao Qiu (mhqiu@stanford.edu)

Received: 24 March 2022 – Discussion started: 1 April 2022

Revised: 24 June 2022 – Accepted: 9 July 2022 – Published: 19 August 2022

Abstract. Evaluating the influence of anthropogenic-emission changes on air quality requires accounting for the influence of meteorological variability. Statistical methods such as multiple linear regression (MLR) models with basic meteorological variables are often used to remove meteorological variability and estimate trends in measured pollutant concentrations attributable to emission changes. However, the ability of these widely used statistical approaches to correct for meteorological variability remains unknown, limiting their usefulness in the real-world policy evaluations. Here, we quantify the performance of MLR and other quantitative methods using simulations from a chemical transport model, GEOS-Chem, as a synthetic dataset. Focusing on the impacts of anthropogenic-emission changes in the US (2011 to 2017) and China (2013 to 2017) on PM_{2.5} and O₃, we show that widely used regression methods do not perform well in correcting for meteorological variability and identifying long-term trends in ambient pollution related to changes in emissions. The estimation errors, characterized as the differences between meteorology-corrected trends and emission-driven trends under constant meteorology scenarios, can be reduced by 30 %–42 % using a random forest model that incorporates both local- and regional-scale meteorological features. We further design a correction method based on GEOS-Chem simulations with constant-emission input and quantify the degree to which anthropogenic emissions and meteorological influences are inseparable, due to their process-based interactions. We conclude by providing recommendations for evaluating the impacts of anthropogenic-emission changes on air quality using statistical approaches.

1 Introduction

Researchers and policymakers have long been interested in understanding the anthropogenic drivers of trends in observed air pollutant concentrations in order to inform air quality policies. Declining trends in pollutant concentrations such as particulate matter with diameters less than 2.5 μm (PM_{2.5}) have been observed in many countries that adopted policies to limit anthropogenic emissions such as SO₂ and NO_x, including the US (McClure and Jaffe, 2018)

and China (Zhang et al., 2019). As the information on anthropogenic emissions is often unavailable or very uncertain, researchers and policymakers often rely on the trends in measured air pollutants to assess the effects of policies. Attributing trends in observed concentrations to anthropogenic-emission changes requires correcting for the influence of changing meteorology, which has become increasingly important but challenging in a changing climate (Saari et al., 2019). Numerous papers attempt to use statistical methods to separate impacts of meteorology from emission changes in

evaluating trends in air quality, but the performance of these commonly used statistical approaches remains unassessed. Further, the impacts of meteorological variability may not even be distinguishable from air quality trends driven by anthropogenic-emission changes, due to their interactions; the magnitude of this interaction also remains unquantified. In this paper, we devise a model-based experiment for evaluating the performance of different statistical methods used for meteorological corrections. We focus on a case of identifying emission-driven linear trends in measured concentrations of PM_{2.5} and ozone (O₃), when information on the anthropogenic emission is not available.

Measured pollutant concentrations are often used as the primary basis for evaluating air quality actions. For example, in 2013, China's central government established targets that aimed to reduce annual average PM_{2.5} concentrations of three urban clusters by 15 % to 25 % between 2012 and 2017 (State Council of the People's Republic of China, 2013). This later translated into a stringent and binding target of a maximum annual mean PM_{2.5} concentration of 60 µg m⁻³ in 2017 for Beijing, which was ultimately reached (the 2017 concentration was 58.5 µg m⁻³) (Beijing Municipal Ecology and Environment Bureau, 2013). However, several studies estimated that the concentration would have exceeded this target in Beijing were it not for meteorological conditions in the winter 2017 that favored pollution reductions (Vu et al., 2019; Chen et al., 2019; Cheng et al., 2019). The European Union and US Environmental Protection Agency (EPA) use a 3-year average of the PM_{2.5} concentration to determine compliance with air quality standards (European Union, 2020; US Environmental Protection Agency, 2019). The US EPA has also proposed to use statistical approaches that aim to correct for the impacts of weather variability on O₃ concentrations in the designation processes (Wells et al., 2021).

Many studies use multiple linear regression (MLR) models with basic meteorological variables to correct for meteorological variability in order to estimate the impacts of emission changes on measured air quality (Otero et al., 2018; Zhai et al., 2019; Li et al., 2018, 2020; Han et al., 2020; Chen et al., 2020). Zhai et al. (2019) and Li et al. (2020) use MLR models to estimate the degree to which trends in PM_{2.5} and O₃ from 2013 to 2019 in China were driven by anthropogenic-emission changes. They first use MLR to predict the PM_{2.5} and O₃ concentrations with meteorological variables and then interpret the residuals of the MLR model as signals resulting from emission changes. A related approach is to combine MLR with techniques that can decompose time series of observed concentrations into long-term, seasonal, and short-term components (e.g., Kolmogorov–Zurbenko (KZ) filters, Zurbenko, 1994). Ma et al. (2016) and Chen et al. (2019) use KZ filters to calculate the long-term component of observed PM_{2.5} and then apply MLR to separate the impacts of long-term meteorological changes on the concentrations. Henneman et al. (2015) apply MLR to the short-term component (identified by KZ filters) of air pollu-

tant concentrations near Atlanta during 2000 to 2012 to separate the impact of short-term meteorological variability and then estimate the long-term trend in air quality.

Other statistical methods including non-linear regression or machine learning models have also been used to correct for meteorological variability (Holland et al., 1998; Carslaw et al., 2007; Hayn et al., 2009; Vu et al., 2019). One popular method is to use a generalized additive model (GAM) to estimate non-linear smooth functions of each meteorological variable within a given smoothing-function family with penalization on non-smoothness. The US EPA uses a GAM model of temperature, wind direction and speed, humidity, pressure, stability, transport trajectories, and synoptic weather to perform weather corrections in assessing long-term trends in O₃ (Camalier et al., 2007). An increasing number of studies use machine learning models (Grange et al., 2018; Vu et al., 2019; Zhang et al., 2020; Shi et al., 2021; Qu et al., 2020). Vu et al. (2019) use a random forest model to predict pollutant concentrations in Beijing with time index and meteorological variables and then calculate the “weather-normalized” concentration for each day with 1000 sets of meteorological fields drawn from the historical meteorological data. They found that the decrease in PM_{2.5} during 2013 to 2017 was largely driven by emission reductions, although the magnitude of reduction is smaller when correcting for meteorological variability.

Despite a large number of papers that apply various meteorology correction methods, very little is known about whether these methods can effectively correct for meteorological variability and thus realistically estimate the counterfactual air quality and reveal the underlying impacts of anthropogenic-emission changes. Most studies cite the prediction performance of their statistical models (such as R^2 and/or mean squared errors) to justify their method choice and analysis. However, good prediction performance does not guarantee the correct estimation of counterfactuals and causal effects (Runge et al., 2019). The performance of these meteorology-corrected methods is unable to be assessed using observational data alone, as the underlying emission-driven trends without influence from meteorological variability cannot be derived from data. Further, statistical analyses often assume that the influence of meteorological variability on pollutant concentration can be cleanly separated from the influence of anthropogenic-emission changes. This is not completely possible, as the impacts of meteorological variability on pollutant concentration will also vary depending on the emissions. The degree to which this interaction affects the ability to calculate emission-related trends under changing meteorology also remains unknown.

Here, we conduct a model experiment to evaluate the performance of widely used statistical models in correcting for meteorological variability and estimating emission-driven trends in air quality (see Fig. 1). We focus on the impacts of anthropogenic-emission changes on annual PM_{2.5} and summer O₃ in the US (2011–2017) and China (2013–

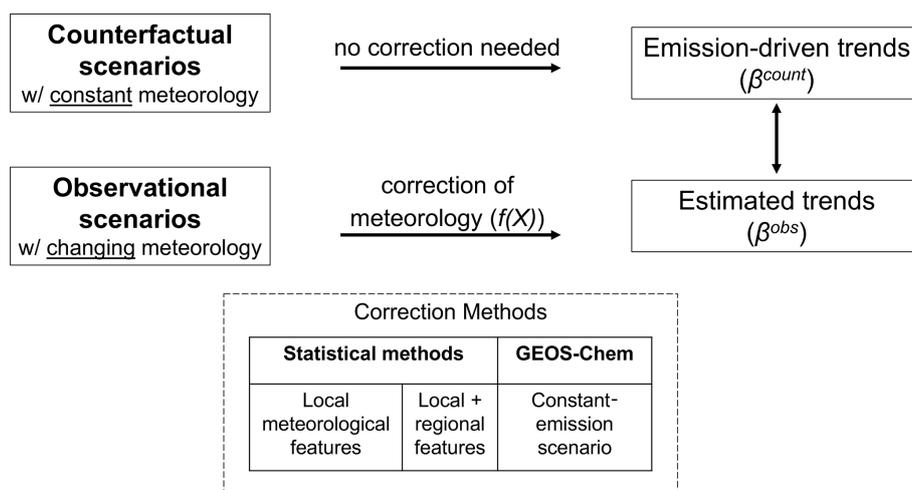


Figure 1. Overview of research methodology. Terms and coefficients are linked to the associated terms in Eq. (1) and Table 1.

2017), two periods well-studied in previous literature. Using a three-dimensional atmospheric chemical transport model, GEOS-Chem, we simulate two sets of scenarios – “observational scenarios” with assimilated meteorological inputs (with interannual variability) and “counterfactual scenarios” with constant meteorological inputs. Using simulated daily concentrations in the observational scenarios, we estimate meteorology-corrected trends for each grid cell from regression models using different statistical correction methods. We then compare the derived trends with the emission-driven trends in the counterfactual scenarios (which are free of meteorological variability by design), calculating the resulting “error” in trend estimation. We further design a correction method based on GEOS-Chem constant-emission simulations and use it to quantify the degree to which attribution to meteorology and emissions separately is possible. Finally, we apply the different statistical correction methods to observational data from surface monitoring networks in the US and China, discussing the variability across different methods. We conclude by providing recommendations for techniques to evaluate air pollution policies under changing meteorological conditions.

2 Method

2.1 GEOS-Chem

GEOS-Chem is a global three-dimensional chemical transport model driven by assimilated meteorological data from the Goddard Earth Observation System (GEOS-5) of the NASA Global Modeling and Assimilation Office (GMAO) (Bey et al., 2001; <http://www.geos-chem.org/>, last access: March 2022). The simulation of $\text{PM}_{2.5}$ in GEOS-Chem represents an external mixture of secondary inorganic aerosols, carbonaceous aerosols, sea salt, and dust aerosols. GEOS-Chem includes detailed O_3 – NO_x –volatile organic carbon

(VOC)–aerosol–halogen tropospheric chemistry (Travis et al., 2016; Sherwen et al., 2016). The GEOS-Chem model has been previously used to study the changes in $\text{PM}_{2.5}$ and O_3 during our studied periods, and model simulations have been shown to be consistent with the observed concentrations (see, e.g., C. Li et al., 2017; Xie et al., 2019, for the US, and Li et al., 2018; Lu et al., 2019; Zhai et al., 2021, for China). Studies in both regions show that the GEOS-Chem model is able to reproduce the spatial, seasonal, and interannual variability and the long-term trends in observed pollutant concentrations, despite biases in absolute concentrations in certain species and regions (Heald et al., 2012; Travis et al., 2016; Tian et al., 2021).

We use GEOS-Chem version 12.3.0 with a horizontal resolution of $0.5^\circ \times 0.625^\circ$ in North America and Asia (Wang et al., 2004). For each scenario, we first conduct a global run at a horizontal resolution of $4^\circ \times 5^\circ$, with a 12-month spin-up. These global runs are then used as the boundary conditions for nested simulations in the US and China with finer resolution of $0.5^\circ \times 0.625^\circ$.

2.2 GEOS-Chem scenarios

Table 1 shows the simulations included in our model experiments. We simulate two sets of scenarios – observational scenarios with interannual variability in meteorology and counterfactual scenarios with constant meteorological inputs. Both scenarios use the same emission inventory as input (see Sect. 2.3). For each grid cell, we estimate the linear trends in pollutant concentrations from simulated daily $\text{PM}_{2.5}$ and O_3 concentrations. We focus on the daily 24 h average $\text{PM}_{2.5}$ over all seasons and the maximum daily average 8 h (MDA8) O_3 in summer (June, July, August). Our focus on the 3 summer months is consistent with many previous studies (e.g., Shen et al., 2015), although this may not capture the peak ozone season for certain regions of the US

Table 1. Overview of GEOS-Chem scenarios and meteorological correction methods. RF: random forest. LASSO: least absolute shrinkage and selection operator.

GEOS-Chem scenarios	Emissions inventory	Meteorological fields	Trend estimates	Meteorological correction
Counterfactual scenarios	<i>Changing</i> 2011–2017 (US) 2013–2017 (China)	<i>Constant</i> 2011 (US) 2013 (China)	β^{count}	
Observational scenarios	<i>Changing</i> 2011–2017 (US) 2013–2017 (China)	<i>Changing</i> 2011–2017 (US) 2013–2017 (China)	$\beta^{\text{uncorrected}}$ β^{MLR} β^{GAM} β^{RF} $\beta^{\text{LASSO-regional}}$ $\beta^{\text{RF-regional}}$ β^{gc}	No correction MLR using local features GAM using local features RF using local features LASSO using local and regional features RF using local and regional features Simulations from constant-emission scenarios
Constant-emission scenarios	<i>Constant</i> 2011 (US) 2013 (China)	<i>Changing</i> 2011–2017 (US) 2013–2017 (China)		

and China. Our GEOS-Chem simulations use meteorological fields from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro et al., 2017). We aggregate the hourly meteorological data for consistency with the pollutant concentrations: a 24 h average for $\text{PM}_{2.5}$ analysis and the corresponding 8 h average for O_3 . Meteorological features that are used in the statistical models can be found in Sect. 2.4.

2.2.1 Observational scenarios

Observational scenarios simulate $\text{PM}_{2.5}$ and O_3 under changing emissions and changing meteorological fields. Trends estimated under the observational scenarios (β^{obs}) are subject to the influences of interannual meteorological variability. Our model experiments were not specifically designed to reproduce observed air quality in these two regions but rather to provide a realistic test case to evaluate the performances of statistical methods. Nevertheless, as shown in Figs. S1 and S2 in the Supplement, the simulated concentrations in $\text{PM}_{2.5}$ and O_3 largely reproduce the daily variability in observed pollutant concentrations. The linear trends in simulated $\text{PM}_{2.5}$ and O_3 concentrations in the observational scenario are largely consistent with trends of the measured concentrations. For example, the average trend (± 1 standard deviation) in the US is $-0.27 \pm 0.30 \mu\text{g}^{-3} \text{yr}^{-1}$ (observation) and $-0.39 \pm 0.24 \text{ppb yr}^{-1}$ (GEOS-Chem) for $\text{PM}_{2.5}$ and $-0.91 \pm 0.98 \text{ppb yr}^{-1}$ (observation) and $-1.02 \pm 0.83 \text{ppb yr}^{-1}$ (GEOS-Chem) for O_3 . The only exception is that our model cannot reproduce the increasing $\text{PM}_{2.5}$ trends in the Northwest US because we do not consider interannual variability in biomass-burning emissions.

2.2.2 Counterfactual scenarios

Counterfactual scenarios simulate $\text{PM}_{2.5}$ and O_3 under changing emissions but constant meteorology. All simulation years in the counterfactual scenario use the meteorological fields of the start year (2011 for the US, 2013 for China). Trends estimated under the counterfactual scenario (β^{count}) are not subject to interannual meteorological variability; we use this as a proxy for the trends in pollutant concentrations driven by emission changes alone. In a sensitivity analysis, we also simulate the counterfactual scenario for China using the meteorological fields for the end year, 2017 (at $4^\circ \times 5^\circ$ resolution, due to computational constraints). We find that the linear trend in $\text{PM}_{2.5}$ and O_3 for each grid cell is highly consistent in the counterfactual scenarios across the choice of the meteorological years (see Fig. S5).

2.2.3 Assumptions for GEOS-Chem experiments

It is important to note that we do not assume our GEOS-Chem simulations perfectly represent the underlying pollutant concentration in the real world (although the model compares relatively well with the observational data). Rather, our main focus is to evaluate how different statistical methods can explain the differences between the observational and counterfactual scenarios. The assumption here is that the differences between observational and counterfactual scenarios are useful approximations of the impacts of meteorological variability on pollutant concentrations. The implications of uncertainty in GEOS-Chem for our results can be found in the “Discussion” section.

2.3 Emission inventory

For the US, we use the National Emissions Inventory 2011 (NEI 2011) as a baseline emission inventory and scale the emissions in 2012 to 2017 to match the annual total emissions each year (US Environmental Protection Agency, 2021b). For China, we use the monthly Multi-resolution Emission Inventory for China (MEIC) during 2013 to 2017 (M. Li et al., 2017; Zheng et al., 2018). During the studied time periods, the US and China experienced dramatic decreases in anthropogenic emissions, particularly in SO₂ and NO_x. In the US, total anthropogenic emissions of SO₂ decreased by 57 % and NO_x emissions decreased by 26 % during 2011 to 2017 (see Fig. S3). In China, anthropogenic SO₂ emissions decreased by 59 %, and NO_x emissions decreased by 21 % during the 2013–2017 period (see Fig. S4).

Natural emissions of multiple chemical species are calculated online in the simulations (rather than prescribed) in the GEOS-Chem model and thus can be influenced by meteorological variability (see Keller et al., 2014, for more details). Impacts of meteorology on PM_{2.5} and O₃ concentrations through changes in the natural emissions are considered here as part of the meteorology–concentration relationship. These emissions include NO_x emissions from lightning and soil processes, sea salt emissions, dust emissions, and biogenic volatile organic carbon (VOC) emissions. However, biomass-burning emissions are prescribed in the GEOS-Chem model, and we hold them constant at the level of the start year. We make this simplification because the GEOS-Chem model uses biomass-burning emissions from external inventories such as the Global Fire Emissions Database (van der Werf et al., 2017), and it is impossible to distinguish natural fire emissions (part of the meteorological variability) from anthropogenic fire emissions (e.g., from farm residual burning). The role of natural emission changes in the meteorology–air quality relationship is further expanded on in the “Discussion” section.

2.4 Statistical and machine learning models

2.4.1 Model with local meteorological variables

We assess the performance of statistical and machine learning models to correct for the meteorological variability in the observational scenarios. We evaluate these methods with a commonly used framework (e.g., used in Li et al., 2018, and Zhai et al., 2019) which models the air pollutant concentrations of each individual grid cell using an additive form of a trend component, a meteorology component, and time fixed effects (to capture daily and monthly variability not related to meteorology). More specifically, we estimate the following regression equation for each grid cell i :

$$y_{it} = \beta_i^{\text{obs}} \times t + f_i(X_{it}) + \eta_{it} + \epsilon_{it}, \quad (1)$$

where y_{it} denotes the PM_{2.5} or O₃ concentration at grid cell i on day t . t is the time index (e.g., in the US, $t = 1$ for 1

January 2011 and $t = 2$ for 2 January 2011). X_{it} denotes the local meteorology features (i.e., meteorological variables in grid cell i on day t). η_{it} is the month-of-year \times day-of-month fixed effect to capture daily and monthly variability in pollutant concentrations that are not related to the meteorological variability (e.g., seasonal cycle in O₃ and PM_{2.5}). ϵ_{it} is the normally distributed error term. β_i^{obs} represents the meteorology-corrected trend in PM_{2.5} or O₃ concentration for grid cell i estimated with the standard ordinary least-squares method. We use the absolute differences $|\beta_i^{\text{obs}} - \beta_i^{\text{count}}|$ to evaluate the performance of different methods to correct for meteorological variability for any given grid cell i .

Here, $f_i(X_{it})$ represents the specifications of local meteorological features for grid cell i under different methods. In addition to the commonly used multiple linear regression (MLR) model, we also evaluate the following models with higher flexibility: polynomial regression models (quadratic, cubic), cubic spline models, generalized additive models (GAM, implemented with R package “mgcv” by Wood, 2011), and random forest (RF) models. We refer to the trend estimates estimated without $f_i(X_{it})$ as uncorrected. We focus on the methods in Table 1 in the main text, and the performance of the other methods can be found in Tables S1 and S2 in the Supplement. Note that the time fixed effects are modeled differently in RF models due to the estimation procedure. More details on the implementation of RF can be found in the Supplement.

We use the following 10 variables from MERRA-2 as our selected meteorological features for the statistical analysis: surface temperature, precipitation, humidity, planetary boundary layer height, cloud fraction, surface air pressure, and wind speed (U and V direction, at surface and 850 hPa levels). These variables are the most commonly used features in previous studies. We also perform sensitivity analyses that include nine more meteorological features: direct photosynthetically active radiation, diffuse photosynthetically active radiation, tropopause pressure, friction velocity, topsoil moisture, root soil moisture, snow depth, surface albedo, and surface air density. These features are selected because they are used as primary or intermediate inputs for calculating PM_{2.5} or O₃ concentrations in the GEOS-Chem model and may be relevant for the variability in pollutant concentrations.

2.4.2 Model with local and regional meteorological variables

We also evaluate models that use both local and regional meteorological features. Regional meteorological features are important for explaining variability in local pollutant concentrations due to (1) pollution transport from neighboring locations and (2) influences from meteorological systems at the synoptic scale (i.e., large-scale weather systems that span over 1000 km such as circulation patterns) (Tai et al., 2012;

Shen et al., 2015; Zhang et al., 2018; Leung et al., 2018; Han et al., 2020). As the incorporation of both local and regional features can quickly expand the dimensionality of the feature space, here we use the least absolute shrinkage and selection operator (LASSO) and the random forest (RF) model, two statistical models that show good prediction performances with high-dimensional data inputs. We estimate the following equations:

$$y_{it} = \beta_i^{\text{obs}} \times t + g_i(X_{it}, Z_t) + \eta_{it} + \epsilon_{it}, \quad (2)$$

where $g_i()$ denotes the functional form fitted by LASSO or RF. X_{it} again denotes the local meteorology features for grid cell i on day t . Z_t denotes the regional-scale meteorology features including the meteorological features for every grid cell in the US on day t (98 cells in $4^\circ \times 5^\circ$; we choose a relatively coarse resolution due to computational cost). Meteorological information in each location in the US may help explain the pollutant concentrations in grid cell i . In total, we have 10 local features (X_{it}) and $10 \times 98 = 980$ regional-scale features (Z_t). The coefficient β_i^{obs} is obtained with the double machine learning approach by Chernozhukov et al. (2018). In particular, the hyperparameters and coefficients of LASSO and RF are selected and fitted using 4-fold cross-validation to avoid the “overfitting risk”. More details on the implementation of LASSO and RF can be found in the Supplement.

2.5 Correction approach using the GEOS-Chem constant-emission scenario

We further design and evaluate an approach to correct for meteorology variability with GEOS-Chem simulations (referred to as the “constant-emis” approach). The constant-emis approach uses GEOS-Chem simulations with constant anthropogenic emissions and changing meteorological fields (“constant-emission scenarios” in Table 1). All years in the constant-emission scenario use anthropogenic emissions of the start year (2011 for the US, 2013 for China). Note that the constant-emis approach also aims to estimate meteorology-corrected trends from the simulated concentrations in the observational scenario. We estimate the following equations:

$$y_{it} = \beta_i^{\text{gc}} \times t + \text{SIM}_{it} + \eta_{it} + \epsilon_{it}, \quad (3)$$

where y_{it} denotes the simulated concentrations in the observational scenario, SIM_{it} denotes the simulated concentrations on day t in grid cell i in the constant-emission scenarios. SIM_{it} serves a similar purpose as the term “ $f_i(X_{it})$ ” in Eq. (1) but comes from the GEOS-Chem simulation. Some previous studies have also used model simulations with constant-emission input as a way to characterize meteorological variability (Zhong et al., 2018; Zhao et al., 2020). β_i^{gc} is the estimated meteorology-corrected trend for $\text{PM}_{2.5}$ or O_3 concentration using this model-based correction method.

Compared to previous statistical and machine learning approaches, the constant-emis approach better captures the

meteorological variability as simulated in GEOS-Chem (as SIM_{it} is directly taken from GEOS-Chem). Therefore, the difference between the trend estimates (β^{gc}) and counterfactual trends (β^{count}) provides a conceptual minimum for estimation errors using the framework of Eq. (1) to perform meteorological corrections. The commonly used framework of Eq. (1) assumes that the impacts of meteorology variability can be separated from the impacts of anthropogenic emissions. In our experiments, this assumption indicates that the differences between the counterfactual scenario and the observational scenario can be solely explained by the meteorological variables. However, the difference in pollutant concentrations between these scenarios is also in part driven by emissions in their interaction with meteorology (despite the fact that our different scenarios use the same emission inventory). We use $|\beta_i^{\text{gc}} - \beta_i^{\text{count}}|$ to quantify the estimation error associated with ignoring such interactions in this framework.

2.6 Air quality observation data

We use the surface air quality measurements from the Air Quality System administered by the US EPA (US Environmental Protection Agency, 2021a). We use the daily 24 h average of $\text{PM}_{2.5}$ concentrations for all months and the daily maximum 8 h average (MDA8) O_3 concentrations for June, July, and August. Figure S1 shows the locations, trends in measured concentrations, and correlations between GEOS-Chem simulations and measured concentrations.

The surface air quality measurements in China are derived from the monitoring network administered by China’s Ministry of Ecology and Environment (2021). The monitoring network was launched in 2013 and has expanded to all prefecture-level cities in mainland China. We use the daily 24 h average of $\text{PM}_{2.5}$ concentrations and the MDA8 O_3 concentrations for summer. Figure S2 shows the locations, trends in measured concentrations, and correlations between GEOS-Chem simulations and measured concentrations.

We use the meteorological variables from MERRA-2 when performing meteorology corrections at these monitoring stations because the meteorology information is not available for all these variables at the station level. This is consistent with previous analyses estimating the meteorology-corrected trends using observational air quality data (e.g., Li et al., 2018).

3 Results

3.1 Performance of different correction methods: US (2011–2017)

Figure 2a and c show the trends in $\text{PM}_{2.5}$ and O_3 concentrations in the counterfactual scenarios in the US. When holding meteorological fields constant across years, decreasing trends in the simulated $\text{PM}_{2.5}$ concentrations across the US result from decreasing anthropogenic emissions. The coun-

terfactual scenario also shows negative linear trends in O₃ concentrations in all but three grid cells in the western US. Increases in summer O₃ in these locations result from the non-linear relationship between O₃ concentrations and NO_x emissions.

Figure 2b and d show the degree to which different meteorological correction methods can recover the emission-driven trends in the counterfactual scenarios. When no correction for meteorology is performed (“uncorrected” in Fig. 2b), we observe large estimation errors in trend estimates over the northeastern and southern US by up to $0.25 \mu\text{g m}^{-3} \text{yr}^{-1}$, an error that is 50 % of the counterfactual trends. We find that the widely used MLR method does not help reduce these errors in PM_{2.5} trend attributions. MLR has a modest impact on reducing the errors in the northeastern US, but it does not decrease the errors over the southern US and leads to even higher errors over the midwestern US. Nationwide, the average magnitude of errors (relative to the counterfactual scenario) *increases* with the MLR correction ($0.083 \mu\text{g m}^{-3} \text{yr}^{-1}$) compared to the uncorrected case ($0.066 \mu\text{g m}^{-3} \text{yr}^{-1}$). Among the five methods, we find that the RF model using both local- and regional-scale features (“RF-regional” in Fig. 2) offers the best performance in recovering the trends in the counterfactual scenarios and is the only method that yields smaller errors than the uncorrected case (the nationwide average error decreased by $0.019 \mu\text{g m}^{-3} \text{yr}^{-1}$; 28 % less). The RF-regional model also outperforms the RF-local and LASSO-regional models, suggesting the importance of considering non-linearity, interactions between different meteorological features, and regional meteorology information in correctly adjusting for the impacts of meteorology.

Meteorological variability also has a substantial influence on the summertime O₃ trends in the US during this period (as shown in Fig. 2d). Relative to the counterfactual scenario, the uncorrected O₃ trends are biased by over $1\text{--}2 \text{ ppb yr}^{-1}$ in large areas of California, the Midwest and South US (as much as 320 % of the counterfactual trends). This is largely driven by the fact that the 2011 and 2012 summers were particularly hot in these regions and led to higher concentrations of O₃ at the beginning of this 7-year period (see Fig. S7 for the South and Midwest US). Therefore, failure to correct for meteorological variability results in much more negative trend estimates in the O₃ concentrations in these areas compared to the counterfactual scenario (see Fig. S6). Meteorology corrections with MLR or GAM help reduce these estimation errors substantially (nationwide average error is reduced by 51 % using MLR or 57 % using GAM compared to uncorrected trends), while large errors still persist in the Midwest and South US. Similar to the case of PM_{2.5}, the RF-regional model offers the best performance in correcting for meteorological variability (the national average error is further reduced by 42 %, compared to MLR), and it is especially helpful in reducing the errors over the Midwest and South US

(regional average error is reduced by 64 % and 44 %, respectively, compared to MLR).

3.2 Performance of different correction methods: China (2013–2017)

Figure 3a and c show the trends in PM_{2.5} and O₃ concentrations in the counterfactual scenarios in China. We find a substantial decline in simulated PM_{2.5} concentration during 2013 to 2017, particularly in eastern and central China. In contrast, there is little change in the simulated PM_{2.5} concentrations in western China in the counterfactual scenario, where PM_{2.5} is dominated by dust species largely driven by natural processes (see Fig. S9). For summer O₃, there are decreasing trends in the counterfactual scenario in most parts of China, except for northern China and some urban areas. This is largely consistent with previous studies that attempt to attribute emission-related changes in O₃ concentrations during this period based on modeling or observational data (Li et al., 2018, 2020; Lu et al., 2020).

Figure 3b shows the magnitude of estimation errors in the trend estimates of annual PM_{2.5} in China under different correction methods. We find the underlying meteorological variability has a substantial impact on PM_{2.5} trends in China during this period. We observe large differences between the uncorrected and counterfactual trends in simulated PM_{2.5} concentrations, particularly in Northwest and Northeast China. Similar to the model experiments in the US, we find that MLR and GAM methods fail to correct for this underlying meteorological variability and lead to further increases in estimation errors in many locations. Relative to the counterfactual scenario, the nationwide average error increases to $0.90 \mu\text{g m}^{-3} \text{yr}^{-1}$ with MLR and $1.06 \mu\text{g m}^{-3} \text{yr}^{-1}$ with GAM (compared to $0.89 \mu\text{g m}^{-3} \text{yr}^{-1}$ with no correction). We find that the RF-regional model recovers the counterfactual trends better than other methods (nationwide average error: $0.64 \mu\text{g m}^{-3} \text{yr}^{-1}$; an improvement of 30 % relative to MLR), but it is still not able to correct for the persistent estimation errors over Northwest China. We further analyze the performance of correction methods for the different component species of PM_{2.5}. As shown in Figs. S10 and S11, the MLR model is particularly unable to correct for the impacts of meteorological variability on nitrate and dust species. Compared with MLR, the RF-regional model better corrects for the impacts of meteorology on secondary organic aerosol species in southern and central China and ammonium in Northeast China but only yields modest improvement in correcting for the errors in dust concentrations over Northwest China (see Fig. S12). In a sensitivity analysis, we use an approach that first fits RF-regional models of each individual PM_{2.5} species and then combines predictions for each species to derive trend estimates. The results are largely similar to the main approach that directly fits the total PM_{2.5} concentration (see Fig. S13).

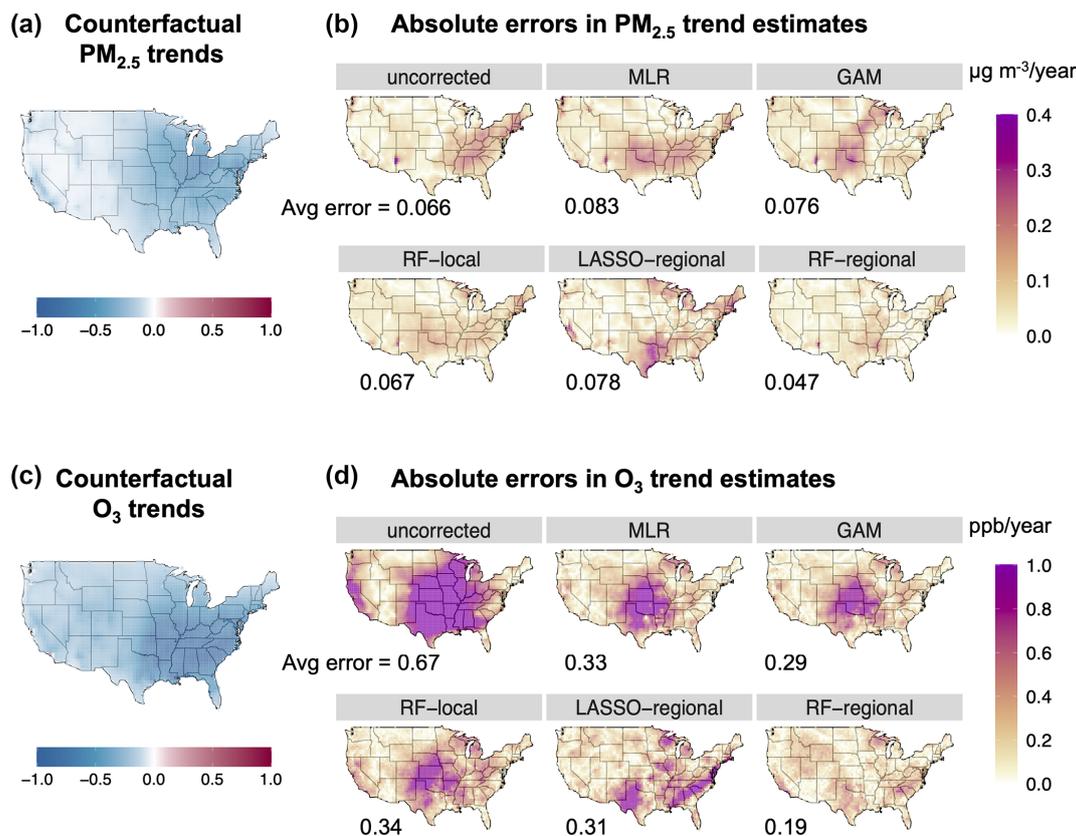


Figure 2. Trend estimates of daily annual $\text{PM}_{2.5}$ (a, b) and summer O_3 (c, d) in the US. Panels (a) and (c) show trend estimates under the counterfactual scenario (β^{count}). Panels (b) and (d) show the absolute magnitude of errors in trend estimates under different correction methods compared with the counterfactual scenarios ($|\beta^{\text{obs}} - \beta^{\text{count}}|$). The average of the absolute errors for each method is shown in the figure (unit of trend estimate: $\mu\text{g m}^{-3} \text{yr}^{-1}$ for $\text{PM}_{2.5}$ and ppb yr^{-1} for O_3).

Figure 3d shows the magnitude of errors in the trend estimates for summer O_3 under different correction methods in China. We find that the MLR model only modestly reduces the estimation errors compared to the uncorrected cases, and the RF-regional model offers the best overall performance. The nationwide average error is reduced to 0.28 ppb yr^{-1} using the RF-regional model (relative to 0.43 ppb yr^{-1} uncorrected and 0.41 ppb yr^{-1} with MLR). Similar to the evaluation of summer time O_3 in the US, we find the non-linear models (GAM, RF-local) perform better than MLR but are not as good as the RF-regional model. Surprisingly, the LASSO-regional model performs the worst in recovering the counterfactual trends. Compared to the US case, we find that the impacts of meteorological variability on O_3 and the method performances are much more spatially heterogeneous (see Figs. S6 and S8), which may be partially due to the more heterogeneous O_3 regimes in China during this period.

3.3 Limitations in separating meteorological and emission influence: quantified with constant-emission scenarios

In our model experiments in both the US and China, we find large differences remain between the trends evaluated with statistical models (even the best-performing RF-regional model) and counterfactual trends. The remaining differences could result from two different factors: (1) the statistical model cannot capture the complex relationship between meteorology and pollutant concentrations, and/or (2) the differences between the observational scenarios and counterfactual scenarios depend on not only the meteorological variability but also the anthropogenic emissions in their interaction with meteorology (i.e., impacts of meteorology on air quality also depend on the level of emissions).

We quantify the potential magnitude of this second factor using our constant-emis approach. As the constant-emis approach captures the exact relationship between meteorology and pollutant concentrations in GEOS-Chem, the error in the constant-emis approach is only associated with the second factor above and thus provides a conceptual minimum of the estimation errors that can be achievable by any statistical

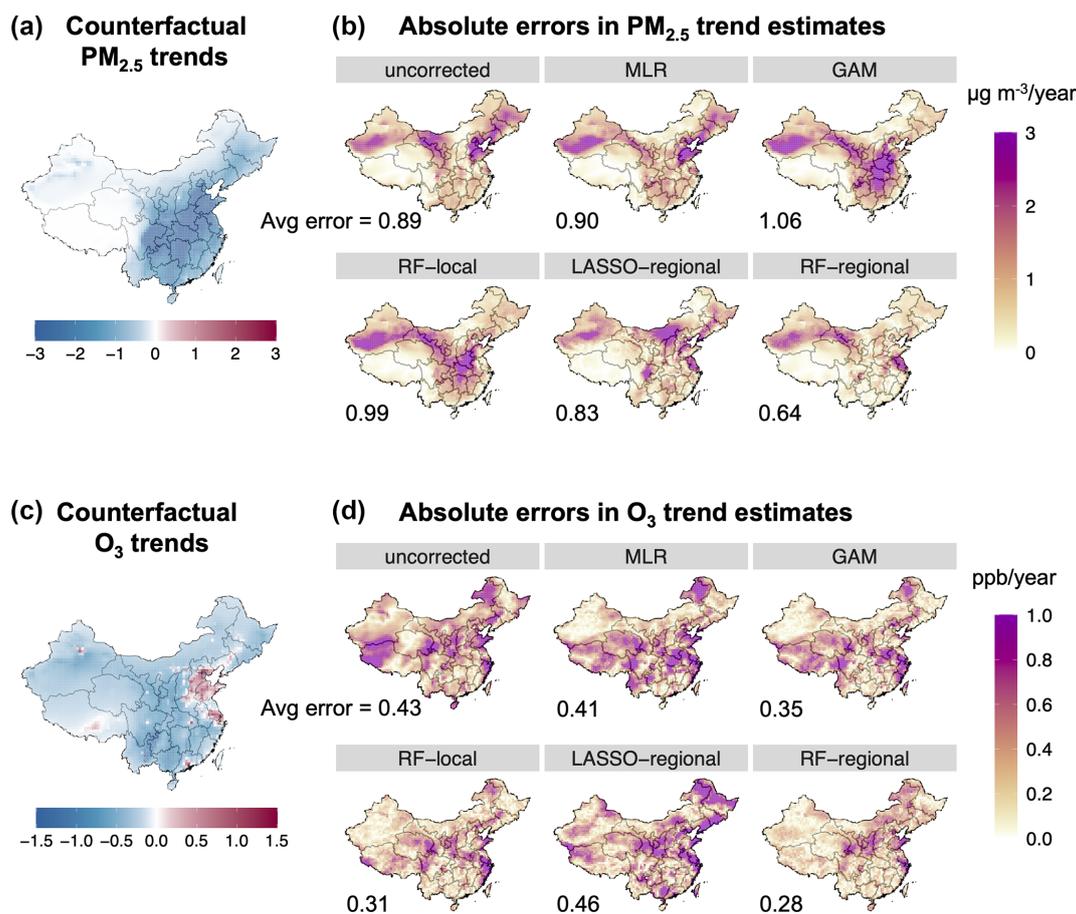


Figure 3. Trend estimates of daily annual $\text{PM}_{2.5}$ (a, b) and summer O_3 (c, d) in China. Panels (a) and (c) show trend estimates under the counterfactual scenario (β^{countf}). Panels (b) and (d) show the absolute magnitude of errors in trend estimates under different correction methods compared with the counterfactual scenarios ($|\beta^{\text{obs}} - \beta^{\text{countf}}|$). The average of the absolute errors for each method is shown in the figure (unit of trend estimate: $\mu\text{g m}^{-3} \text{yr}^{-1}$ for $\text{PM}_{2.5}$ and ppb yr^{-1} for O_3).

approach. Figure 4 shows the estimation errors in trend estimates using the constant-emission scenarios simulated by GEOS-Chem. We focus on the trends in summer O_3 in the US and annual $\text{PM}_{2.5}$ in China, for which we see the largest impacts of meteorological variability on the pollutant trends and the largest improvements in reducing estimation errors from the correction methods. Compared to the statistical models (e.g., MLR and RF-regional in Fig. 4a and c), trends evaluated using the constant-emis approach are very similar to the trends in the counterfactual scenarios. The national average error in trend estimates is only 0.04 ppb yr^{-1} for the O_3 trends in the US (relative to 0.33 ppb yr^{-1} under MLR or 0.19 ppb yr^{-1} under RF-regional) and only $0.08 \mu\text{g m}^{-3} \text{yr}^{-1}$ for the $\text{PM}_{2.5}$ trends in China (relative to $0.91 \mu\text{g m}^{-3} \text{yr}^{-1}$ under MLR or $0.64 \mu\text{g m}^{-3} \text{yr}^{-1}$ under RF-regional).

However, the estimation errors calculated above are still non-negligible and can be large in certain regions. As shown in Fig. 4b and d, the constant-emis approach generally yields trend estimates biased by 10 % relative to the counterfactual trends, but the errors can be up to 40 % in certain areas. This

error term is the result of ignoring how emissions could potentially influence the impacts of meteorology on the pollutant concentrations – that is, the impacts of the same meteorological variability on concentrations may be different in the start year (with high emissions) compared to the end year (with low emissions).

3.4 Application to observational data

Figure 5 shows the regional trends in O_3 in the US and trends in $\text{PM}_{2.5}$ in China estimated from the observational data from surface monitoring networks and the GEOS-Chem simulations (only grid cells that overlap with monitor locations are shown here). Here, to correct for the meteorology variability in observational data, we implement the same set of statistical methods as shown in Table 1 but with different numerical coefficients directly derived from the observational data. When applying different meteorological correction methods to the observational data, our analysis reveals that the choice of methods for meteorological correction can yield very dif-

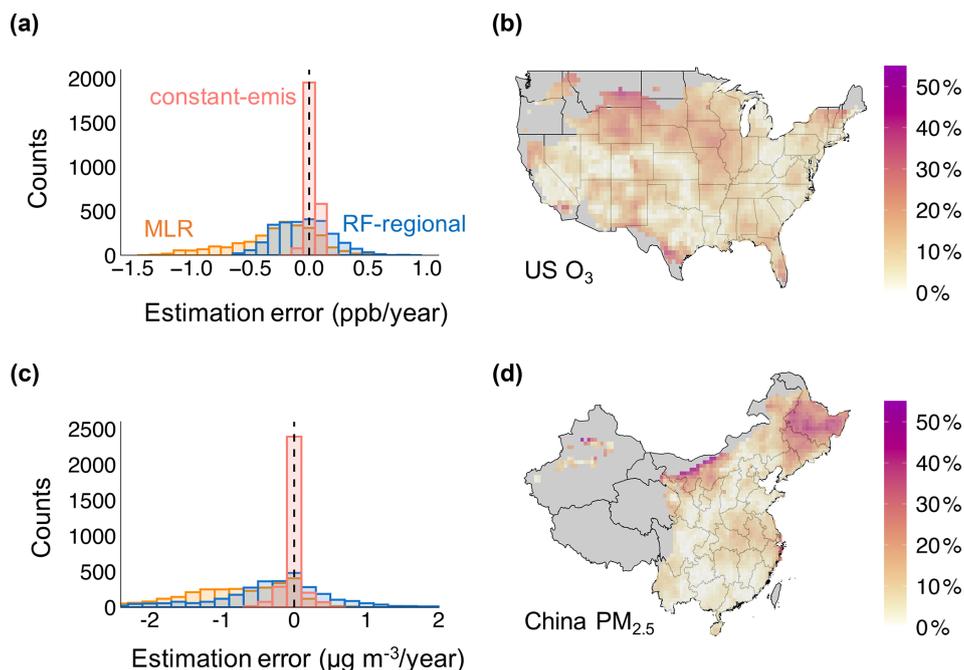


Figure 4. Panels (a) and (c) show the histogram of estimation errors in trend estimates assessed using MLR, RF-regional and constant-emis. Panels (b) and (d) show the errors assessed with the constant-emis method as a percentage of the trends in the counterfactual scenario ($|\beta^{\text{sc}} - \beta^{\text{count}}|/|\beta^{\text{count}}|$). Panels (b) and (d) only show grid cells with a trend in the counterfactual scenarios $> 0.2 \text{ ppb yr}^{-1}$ or $> 0.2 \mu\text{g m}^{-3} \text{ yr}^{-1}$; remaining grid cells are shown in gray. Panels (a) and (b) illustrate the summer O₃ trends in the US. Panels (c) and (d) illustrate the annual PM_{2.5} trends in China.

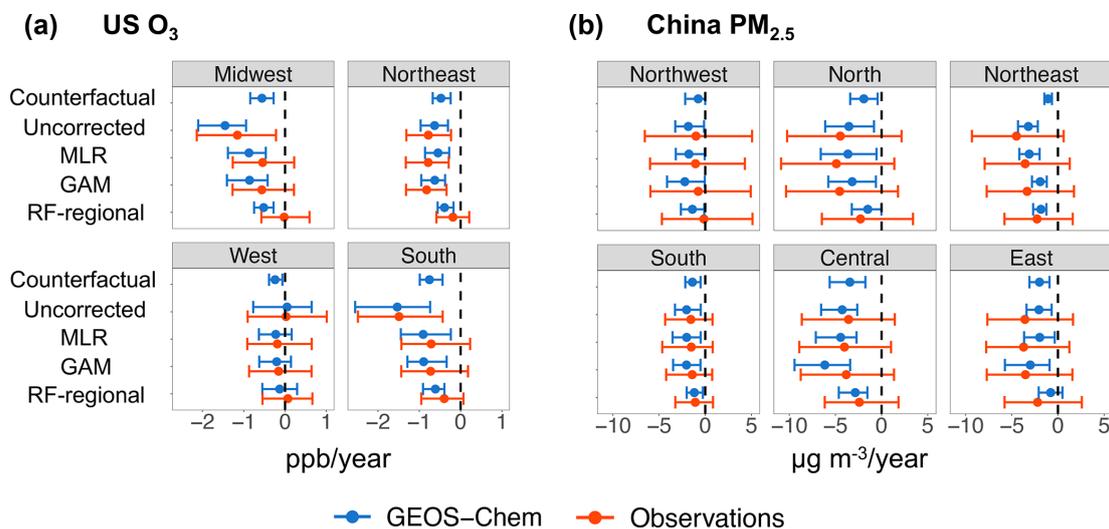


Figure 5. Trends in O₃ in the US (a) and PM_{2.5} in China (b) estimated from the observational data (red) and GEOS-Chem simulations (blue) under different correction methods. Trends in pollutant concentrations are estimated at the monitor level (for the observational data) or at the grid cell level (for GEOS-Chem simulations). The point indicates the average value of the assessed trends of all monitors (or grid cells) within a region. The error bars show the 10th and 90th percentile of the assessed trends of all monitors/grid cells within a region. Panel (a) illustrates the summer O₃ trends in the US (unit: ppb yr⁻¹). Panel (b) illustrates the annual PM_{2.5} trends in China (unit: μg m⁻³ yr⁻¹). We classify the US states into four regions according to the US Census Bureau and classify China's provinces into six regions based on the structure of China's subnational electric grid. Observational data are derived from US Environmental Protection Agency (2021a) and China's Ministry of Ecology and Environment (2021).

ferent results for certain regions. For example, the regional average uncorrected O_3 trend is -1.49 and -1.15 ppbyr^{-1} in the Midwest and South US, respectively, which overestimates the reductions in concentrations attributable to anthropogenic-emission changes (Fig. 5a). Correcting for the meteorological variability with the MLR model yields a regional average trend of -0.54 ppbyr^{-1} in the Midwest (a decrease by 53 % in magnitude relative to uncorrected trends) and -0.71 ppbyr^{-1} in the South US (a decrease by 52 %). RF-regional model further reduces the absolute magnitude of the declines in O_3 attributable to emission reductions to -0.02 ppbyr^{-1} for the Midwest and -0.40 ppbyr^{-1} for the South US. Importantly, these patterns are consistent with the results from our model experiments in these regions: the RF-regional model also estimates a much less negative emission-driven trend in the South US compared to the uncorrected case and MLR estimates in the GEOS-Chem simulations. For the GEOS-Chem simulations, RF-regional estimates are 39 % smaller than MLR estimates, and this is comparable to the magnitude changes for the observational data (RF-regional estimates are 44 % smaller than MLR). As the RF-regional model outperforms the other correction methods in recovering counterfactual trends in the GEOS-Chem simulations, this potentially also suggests a better performance of RF-regional in recovering the underlying emission-driven trends when applying to the observational data.

We find similar consistency in the method performances between observational data and GEOS-Chem simulations in China as well (Fig. 5b). When applying to the observational data from the surface monitoring network, a much smaller reduction in $\text{PM}_{2.5}$ concentrations is attributed to anthropogenic-emission changes in North, Northeast and East China using the RF-regional model, relative to the MLR estimates. For example, the average emission-driven trend estimated from the observational data is -4.9 $\mu\text{g m}^{-3} \text{yr}^{-1}$ in Beijing under the RF-regional model, compared with -9.6 $\mu\text{g m}^{-3} \text{yr}^{-1}$ under the MLR model. These patterns are consistent with the patterns of the trend estimates estimated from our GEOS-Chem simulations with different statistical methods.

4 Discussion

We designed a model experiment that enables us to directly quantify the performance of different statistical models to evaluate the trends in pollutant concentrations driven by anthropogenic-emission changes. Based on our evaluations of either $\text{PM}_{2.5}$ or O_3 trends across the US and China during periods of recent emission declines, our analysis shows that widely used MLR and GAM methods do not perform well in correcting for the meteorological variability and recovering simulated emission-driven trends. We propose a random forest model that uses both local and regional meteorological features, which offers the best overall performance in re-

covering the emission-driven trends across both species and countries. Applying this model to observational data suggests that estimates based on MLR or similar methods may overestimate the impacts of anthropogenic-emission changes on the decline in pollutant concentrations in certain regions in the US and China. However, the RF-regional method does not outperform all the other approaches in every location despite its better overall performance (see Figs. S14 and S15). This suggests that using multiple statistical approaches may be necessary to derive robust conclusions for attributing pollutant trends to emission changes.

With our model experiments, we also quantify the estimation errors in assuming emission impacts can be perfectly separated from meteorological variability. These errors likely bound the estimation errors that can be achieved by any statistical methods with this assumption. In the future, more complex statistical and machine learning methods could be applied to distinguish emission-driven and meteorologically driven changes, but attribution solely based on observed concentrations and meteorology will be limited by physical interactions between emissions and meteorology. We find that the estimation errors resulting from these interactions are overall much smaller compared to the estimation errors in the existing statistical methods but can still be important for certain regions at certain times. However, the intertwined relationships between anthropogenic emissions and meteorology are often much more complex in reality compared to our model experiments. For example, meteorology can also directly influence anthropogenic emissions (e.g., increased electricity consumption during extreme weather conditions, US Energy Information Agency, 2019; He et al., 2020). Therefore, the estimation errors that can be achieved by more flexible statistical models can potentially be even larger than the errors quantified with our constant-emis approach.

While the GEOS-Chem model provides us with a framework to test statistical methods, its use in our model experiments introduces some uncertainty and limitations. Specifically, our experiments assess the performance of statistical methods in correcting for the meteorology–pollution relationships encoded in GEOS-Chem, which may differ from the complex relationships in the observational data. Several studies have shown that GEOS-Chem and similar models do not capture certain meteorology–pollution relationships in the observational data (e.g., temperature– O_3 relationship, Porter and Heald, 2019, and influence of regional meteorological patterns, Fiore et al., 2009). The relationships encoded in GEOS-Chem may be different from the underlying meteorology–pollution relationships in the following three ways: (1) parameters in GEOS-Chem that describe these relationships are uncertain; (2) the relationships in GEOS-Chem are incorrect or incomplete; and (3) the relationships in GEOS-Chem are deterministic compared to the potential stochastic underlying processes. Therefore, the performance of any individual statistical method is likely to be worse in the real world compared to its ability to reproduce a

deterministic meteorology–pollution relationship encoded in GEOS-Chem. Further model-based experiments could apply our methods to different atmospheric models in order to test if these conclusions differ by different models.

Changes in natural emissions due to meteorological variability play an important role in the air quality–meteorology relationship. Our model experiment considers natural emission changes that can be simulated online with assimilated meteorological fields in GEOS-Chem, including soil NO_x emissions, biogenic VOC emissions, and dust emissions. We find that the statistical models perform notably worse in correcting for the variability in dust-related $\text{PM}_{2.5}$ (see Fig. S12 for results using RF-regional), likely because dust $\text{PM}_{2.5}$ is extremely variable, with zero concentration on most non-dust days but extremely high concentration during the occasional dust storms. Our findings can potentially shed light on another important source of natural emissions, wildfire emissions, which are also quite variable but have become an increasingly important contributor to $\text{PM}_{2.5}$ and O_3 in certain regions (e.g., western US) (Burke et al., 2021). While emissions from biomass burning are held constant in our model experiments, as the wildfire emissions are prescribed in GEOS-Chem, wildfire emissions are significantly influenced by climatic variability (Abatzoglou and Williams, 2016; Xie et al., 2022) and will likely be a substantial challenge for any meteorological correction method in the future that attempts to separate changes in anthropogenic emissions from the variability in climate and associated natural emissions.

Our research reveals multiple directions for future research to enhance our understanding of the usage of statistical models to evaluate trends in pollutant concentrations under changing meteorological conditions. One key but challenging question is to better understand the estimation errors in these existing approaches; e.g., why the MLR model is able to correct for the meteorological variability in some locations but not others. In this paper, we only test a selection of methods based on their popularity in the existing literature and propose a simple-to-use model (RF-regional). More complex models (such as convolutional neural networks) may offer better performance, but the estimation error will likely be bounded by the errors in the constant-emis approach. Our work only evaluates the statistical and machine learning models in Eqs. (1) and (2), which only represent one (popular) set of evaluations that performs location-specific trend estimation with adjustments for meteorology and secular trends. However, other statistical model specifications specifically targeted to questions of meteorological interaction or that permit borrowing information across locations may generate different results. Constrained by computational resources and the availability of emission inventories, our simulation only covers a relatively short time period which could result in additional uncertainty in the linear trend estimates. When possible, future studies could evaluate performances of the statistical models with longer simulations

and alternative trend estimates (such as the Theil–Sen estimator). A deeper investigation of the estimation error due to assuming perfect separation between meteorology and emission is also essential for understanding how we should interpret studies that use these statistical methods. For example, further work could explore how these errors will vary by the magnitude of emission reductions and the chemistry regimes.

5 Recommendations for attributing trends to emission changes

Using statistical methods to causally infer relationships between simulated air pollutant concentrations and anthropogenic emissions is challenging, and doing so in contexts of observational data is even more challenging. Understanding the uncertainty in statistical models in characterizing the meteorology–pollution relationship is essential to evaluating the effectiveness of policy interventions with observational data. Here, we make several recommendations to researchers and policymakers based on our analysis.

For those who aim to infer causal effects of emission changes on air quality based on observational data on concentrations and meteorology, we recommend using multiple statistical methods to correct for meteorological variability when evaluating the impacts of policies or interventions on air quality. From our two case studies, we find a relatively large variation between the trend parameters estimated by different statistical methods (especially at the grid cell or monitor level). Some methods perform better in certain locations but not in others (though RF-regional is the best-performing method overall). Using multiple approaches (linear/non-linear and at the local/regional scale) may help to quantify uncertainty related to meteorological corrections. These findings also suggest that empirical analyses may benefit from considering the impacts of meteorological variability on air quality separately for each region or even for each monitor location (if data permit), instead of attempting to determine a general relationship between meteorological variability and air pollution over a large spatial domain. Finally, analysts should be particularly cautious when using statistical methods to estimate impacts of anthropogenic emissions on air quality in regions where pollution variability is dominated by meteorologically influenced environmental processes such as dust emissions, as we consistently show that typical statistical methods (in combination with the standard set of meteorological variables) do not work well in those regions.

Due to the non-negligible estimation errors in recovering the counterfactual trends even with the best-performing statistical approach we test, we believe these statistical analyses are most useful in understanding the patterns of anthropogenic emissions on air quality when aggregated across larger spatial areas, rather than providing specific trends for individual monitor locations. There is a higher degree of con-

sistency among the trend estimates across different methods when aggregated at regional level, but assessment at the local level is more sensitive to method choice. The absolute magnitude of monitor-level trends needs to be interpreted with caution, considering both the uncertainty from the statistical methods and also the limit of meteorological correction due to ignoring the interactions between meteorology and emissions.

Because measured pollutant concentrations are subject to the influence of underlying meteorological variability, many efforts have attempted to correct for the impacts of meteorological variability and use “meteorology-corrected” concentrations and trends to assist in evaluating the effectiveness of air quality policies. Our study evaluates existing methods that aim to correct for the meteorological variability and finds many of these methods do not perform well. This raises potential concerns about the use of meteorology-corrected concentrations as targets for policy evaluation. Meteorology-corrected concentrations and trends remain useful metrics to quantify the influence of emissions. However, a more comprehensive evaluation of the effectiveness of policy requires interpreting measurements with all available tools, ideally including both statistical analyses and physical models.

Code and data availability. The GEOS-Chem simulation of different scenarios and the R scripts to implement the statistical methods to correct for meteorological variability are available at the following repository: <https://doi.org/10.5281/zenodo.6857259> (Qiu et al., 2022). All the other data needed to evaluate the conclusions in the paper are present in the paper.

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/acp-22-10551-2022-supplement>.

Author contributions. MQ and NES designed the research. MQ performed the statistical analysis and GEOS-Chem modeling simulations. All authors interpreted the results and wrote the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We thank Colette Heald and Valerie Karplus for helpful comments and discussions. We thank Yixuan Zheng for assistance with the MEIC emission inventory. We thank Ke Li for sharing code for stepwise MLR analysis.

Minghao Qiu gratefully acknowledges the support of the MIT Martin Family Society of Fellows for Sustainability.

Financial support. This publication was supported by the US EPA (grant no. RD-835872-01). Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA. Further, the US EPA does not endorse the purchase of any commercial products or services mentioned in the publication. This work was also supported by the National Institutes of Health (NIH; grant no. NIEHS R01ES026217).

Review statement. This paper was edited by Anne Perring and reviewed by Benjamin Wells and one anonymous referee.

References

- Abatzoglou, J. T. and Williams, A. P.: Impact of anthropogenic climate change on wildfire across western US forests, *P. Natl. Acad. Sci. USA*, 113, 11770–11775, 2016.
- Beijing Municipal Ecology and Environment Bureau: Beijing Clean Air Action Plan (2013–2017), <http://sthjj.beijing.gov.cn/bjhrb/index/xxgk69/sthjlyzgw/wrygl/603133/index.html> (last access: March 2022), 2013.
- Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *J. Geophys. Res.-Atmos.*, 106, 23073–23095, <https://doi.org/10.1029/2001JD000807>, 2001.
- Burke, M., Driscoll, A., Heft-Neal, S., Xue, J., Burney, J., and Wara, M.: The changing risk and burden of wildfire in the United States, *P. Natl. Acad. Sci. USA*, 118, e2011048118, <https://doi.org/10.1073/pnas.2011048118>, 2021.
- Camalier, L., Cox, W., and Dolwick, P.: The effects of meteorology on ozone in urban areas and their use in assessing ozone trends, *Atmos. Environ.*, 41, 7127–7137, <https://doi.org/10.1016/j.atmosenv.2007.04.061>, 2007.
- Carslaw, D. C., Beevers, S. D., and Tate, J. E.: Modelling and assessing trends in traffic-related emissions using a generalised additive modelling approach, *Atmos. Environ.*, 41, 5289–5299, <https://doi.org/10.1016/j.atmosenv.2007.02.032>, 2007.
- Chen, L., Zhu, J., Liao, H., Yang, Y., and Yue, X.: Meteorological influences on PM_{2.5} and O₃ trends and associated health burden since China’s clean air actions, *Sci. Total Environ.*, 744, 140837, <https://doi.org/10.1016/j.scitotenv.2020.140837>, 2020.
- Chen, Z., Chen, D., Kwan, M.-P., Chen, B., Gao, B., Zhuang, Y., Li, R., and Xu, B.: The control of anthropogenic emissions contributed to 80 % of the decrease in PM_{2.5} concentrations in Beijing from 2013 to 2017, *Atmos. Chem. Phys.*, 19, 13519–13533, <https://doi.org/10.5194/acp-19-13519-2019>, 2019.
- Cheng, J., Su, J., Cui, T., Li, X., Dong, X., Sun, F., Yang, Y., Tong, D., Zheng, Y., Li, Y., Li, J., Zhang, Q., and He, K.: Dominant role of emission reduction in PM_{2.5} air quality improvement in Beijing during 2013–2017: a model-based decomposition analysis, *Atmos. Chem. Phys.*, 19, 6125–6146, <https://doi.org/10.5194/acp-19-6125-2019>, 2019.

- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J.: Double/debiased machine learning for treatment and structural parameters, *Economet. J.*, 21, C1–C68, <https://doi.org/10.1111/ectj.12097>, 2018.
- China's Ministry of Ecology and Environment: National Air Quality Monitoring Data, <https://quotsoft.net/air/>, last access: May 2021.
- European Union: Air Quality Standards in the European Union, <https://ec.europa.eu/environment/air/quality/standards.htm> (last access: March 2022), 2020.
- Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., Doherty, R. M., Horowitz, L. W., MacKenzie, I. A., Sanderson, M. G., Shindell, D. T., Stevenson, D. S., Szopa, S., Van Dingenen, R., Zeng, G., Atherton, C., Bergmann, D., Bey, I., Carmichael, G., Collins, W. J., Duncan, B. N., Faluvegi, G., Folberth, G., Gauss, M., Gong, S., Hauglustaine, D., Holloway, T., Isaksen, I. S. A., Jacob, D. J., Jonson, J. E., Kaminski, J. W., Keating, T. J., Lupu, A., Marmor, E., Montanaro, V., Park, R. J., Pitari, G., Pringle, K. J., Pyle, J. A., Schroeder, S., Vivanco, M. G., Wind, P., Wojcik, G., Wu, S., and Zuber, A.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, *J. Geophys. Res.-Atmos.*, 114, D04301, <https://doi.org/10.1029/2008JD010816>, 2009.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), *J. Climate*, 30, 5419–5454, 2017.
- Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and Hueglin, C.: Random forest meteorological normalisation models for Swiss PM₁₀ trend analysis, *Atmos. Chem. Phys.*, 18, 6223–6239, <https://doi.org/10.5194/acp-18-6223-2018>, 2018.
- Han, H., Liu, J., Shu, L., Wang, T., and Yuan, H.: Local and synoptic meteorological influences on daily variability in summertime surface ozone in eastern China, *Atmos. Chem. Phys.*, 20, 203–222, <https://doi.org/10.5194/acp-20-203-2020>, 2020.
- Hayn, M., Beirle, S., Hamprecht, F. A., Platt, U., Menze, B. H., and Wagner, T.: Analysing spatio-temporal patterns of the global NO₂-distribution retrieved from GOME satellite observations using a generalized additive model, *Atmos. Chem. Phys.*, 9, 6459–6477, <https://doi.org/10.5194/acp-9-6459-2009>, 2009.
- He, P., Liang, J., Qiu, Y. L., Li, Q., and Xing, B.: Increase in domestic electricity consumption from particulate air pollution, *Nature Energy*, 5, 985–995, 2020.
- Heald, C. L., Collett Jr., J. L., Lee, T., Benedict, K. B., Schwandner, F. M., Li, Y., Clarisse, L., Hurtmans, D. R., Van Damme, M., Clerbaux, C., Coheur, P.-F., Philip, S., Martin, R. V., and Pye, H. O. T.: Atmospheric ammonia and particulate inorganic nitrogen over the United States, *Atmos. Chem. Phys.*, 12, 10295–10312, <https://doi.org/10.5194/acp-12-10295-2012>, 2012.
- Henneman, L. R., Holmes, H. A., Mulholland, J. A., and Russell, A. G.: Meteorological detrending of primary and secondary pollutant concentrations: Method application and evaluation using long-term (2000–2012) data in Atlanta, *Atmos. Environ.*, 119, 201–210, <https://doi.org/10.1016/j.atmosenv.2015.08.007>, 2015.
- Holland, D. M., Principe, P. P., and Sickles, J. E.: Trends in atmospheric sulfur and nitrogen species in the eastern United States for 1989–1995, *Atmos. Environ.*, 33, 37–49, [https://doi.org/10.1016/S1352-2310\(98\)00123-X](https://doi.org/10.1016/S1352-2310(98)00123-X), 1998.
- Keller, C. A., Long, M. S., Yantosca, R. M., Da Silva, A. M., Pawson, S., and Jacob, D. J.: HEMCO v1.0: a versatile, ESMF-compliant component for calculating emissions in atmospheric models, *Geosci. Model Dev.*, 7, 1409–1417, <https://doi.org/10.5194/gmd-7-1409-2014>, 2014.
- Leung, D. M., Tai, A. P. K., Mickley, L. J., Moch, J. M., van Donkelaar, A., Shen, L., and Martin, R. V.: Synoptic meteorological modes of variability for fine particulate matter (PM_{2.5}) air quality in major metropolitan regions of China, *Atmos. Chem. Phys.*, 18, 6733–6748, <https://doi.org/10.5194/acp-18-6733-2018>, 2018.
- Li, C., Martin, R. V., Van Donkelaar, A., Boys, B. L., Hammer, M. S., Xu, J. W., Marais, E. A., Reff, A., Strum, M., Ridley, D. A., Crippa, M., Brauer, M., and Zhang, Q.: Trends in Chemical Composition of Global and Regional Population-Weighted Fine Particulate Matter Estimated for 25 Years, *Environ. Sci. Technol.*, 51, 11185–11195, <https://doi.org/10.1021/acs.est.7b02530>, 2017.
- Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., and Bates, K. H.: Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China, *P. Natl. Acad. Sci. USA*, 116, 422–427, <https://doi.org/10.1073/pnas.1812168116>, 2018.
- Li, K., Jacob, D. J., Shen, L., Lu, X., De Smedt, I., and Liao, H.: Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences, *Atmos. Chem. Phys.*, 20, 11423–11433, <https://doi.org/10.5194/acp-20-11423-2020>, 2020.
- Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., and He, K.: Anthropogenic emission inventories in China: A review, *Natl. Sci. Rev.*, 4, 834–866, <https://doi.org/10.1093/nsr/nwx150>, 2017.
- Lu, X., Zhang, L., Chen, Y., Zhou, M., Zheng, B., Li, K., Liu, Y., Lin, J., Fu, T.-M., and Zhang, Q.: Exploring 2016–2017 surface ozone pollution over China: source contributions and meteorological influences, *Atmos. Chem. Phys.*, 19, 8339–8361, <https://doi.org/10.5194/acp-19-8339-2019>, 2019.
- Lu, X., Zhang, L., Wang, X., Gao, M., Li, K., Zhang, Y., Yue, X., and Zhang, Y.: Rapid increases in warm-season surface ozone and resulting health impact in China since 2013, *Environ. Sci. Tech. Lett.*, 7, 240–247, 2020.
- Ma, Z., Xu, J., Quan, W., Zhang, Z., Lin, W., and Xu, X.: Significant increase of surface ozone at a rural site, north of eastern China, *Atmos. Chem. Phys.*, 16, 3969–3977, <https://doi.org/10.5194/acp-16-3969-2016>, 2016.
- McClure, C. D. and Jaffe, D. A.: US particulate matter air quality improves except in wildfire-prone areas, *P. Natl. Acad. Sci. USA*, 115, 7901–7906, <https://doi.org/10.1073/pnas.1804353115>, 2018.
- Otero, N., Sillmann, J., Mar, K. A., Rust, H. W., Solberg, S., Andersson, C., Engardt, M., Bergström, R., Bessagnet, B., Collette, A., Couvidat, F., Cuvelier, C., Tsyro, S., Fagerli, H., Schaap, M., Manders, A., Mircea, M., Briganti, G., Cappelletti, A., Adani, M., D'Isidoro, M., Pay, M.-T., Theobald, M., Vivanco, M. G., Wind, P., Ojha, N., Raffort, V., and Butler, T.: A multi-model comparison of meteorological drivers of sur-

- face ozone over Europe, *Atmos. Chem. Phys.*, 18, 12269–12288, <https://doi.org/10.5194/acp-18-12269-2018>, 2018.
- Porter, W. C. and Heald, C. L.: The mechanisms and meteorological drivers of the summertime ozone–temperature relationship, *Atmos. Chem. Phys.*, 19, 13367–13381, <https://doi.org/10.5194/acp-19-13367-2019>, 2019.
- Qiu, M., Zigler, C., and Selin, N.: Statistical and machine learning methods for evaluating trends in air quality under changing meteorological conditions [Data set], Zenodo [data set and code], <https://doi.org/10.5281/zenodo.6857259>, 2022.
- Qu, L., Liu, S., Ma, L., Zhang, Z., Du, J., Zhou, Y., and Meng, F.: Evaluating the meteorological normalized PM_{2.5} trend (2014–2019) in the “2+26” region of China using an ensemble learning technique, *Environ. Pollut.*, 266, 115346, <https://doi.org/10.1016/j.envpol.2020.115346>, 2020.
- Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., and Zscheischler, J.: Inferring causation from time series in Earth system sciences, *Nat. Commun.*, 10, 1–13, <https://doi.org/10.1038/s41467-019-10105-3>, 2019.
- Saari, R., Mei, Y., Monier, E., and Garcia-Menendez, F.: Effect of Health-related Uncertainty and Natural Variability on Health Impacts and Co-Benefits of Climate Policy, *Environ. Sci. Technol.*, 53, 1098–1108, <https://doi.org/10.1021/acs.est.8b05094>, 2019.
- Shen, L., Mickley, L. J., and Tai, A. P. K.: Influence of synoptic patterns on surface ozone variability over the eastern United States from 1980 to 2012, *Atmos. Chem. Phys.*, 15, 10925–10938, <https://doi.org/10.5194/acp-15-10925-2015>, 2015.
- Sherwen, T., Schmidt, J. A., Evans, M. J., Carpenter, L. J., Großmann, K., Eastham, S. D., Jacob, D. J., Dix, B., Koenig, T. K., Sinreich, R., Ortega, I., Volkamer, R., Saiz-Lopez, A., Prados-Roman, C., Mahajan, A. S., and Ordóñez, C.: Global impacts of tropospheric halogens (Cl, Br, I) on oxidants and composition in GEOS-Chem, *Atmos. Chem. Phys.*, 16, 12239–12271, <https://doi.org/10.5194/acp-16-12239-2016>, 2016.
- Shi, Z., Song, C., Liu, B., Lu, G., Xu, J., Van Vu, T., Elliott, R. J., Li, W., Bloss, W. J., and Harrison, R. M.: Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns, *Sci. Adv.*, 7, eabd6696, <https://doi.org/10.1126/sciadv.abd6696>, 2021.
- State Council of the People’s Republic of China: The Air Pollution Prevention and Control Action Plan (2013–2017), http://www.gov.cn/zwqk/2013-09/12/content_2486773.htm (last access: March 2022), 2013.
- Tai, A. P. K., Mickley, L. J., Jacob, D. J., Leibensperger, E. M., Zhang, L., Fisher, J. A., and Pye, H. O. T.: Meteorological modes of variability for fine particulate matter (PM_{2.5}) air quality in the United States: implications for PM_{2.5} sensitivity to climate change, *Atmos. Chem. Phys.*, 12, 3131–3145, <https://doi.org/10.5194/acp-12-3131-2012>, 2012.
- Tian, R., Ma, X., and Zhao, J.: A revised mineral dust emission scheme in GEOS-Chem: improvements in dust simulations over China, *Atmos. Chem. Phys.*, 21, 4319–4337, <https://doi.org/10.5194/acp-21-4319-2021>, 2021.
- Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., Yu, K., Miller, C. C., Yantosca, R. M., Sulprizio, M. P., Thompson, A. M., Wennberg, P. O., Crouse, J. D., St. Clair, J. M., Cohen, R. C., Laughner, J. L., Dibb, J. E., Hall, S. R., Ullmann, K., Wolfe, G. M., Pollack, I. B., Peischl, J., Neuman, J. A., and Zhou, X.: Why do models overestimate surface ozone in the Southeast United States?, *Atmos. Chem. Phys.*, 16, 13561–13577, <https://doi.org/10.5194/acp-16-13561-2016>, 2016.
- US Energy Information Agency: Heat wave results in highest U. S. electricity demand since 2017, <https://www.eia.gov/todayinenergy/detail.php?id=40253> (last access: March 2022), 2019.
- US Environmental Protection Agency: National Primary and Secondary Ambient Air Quality Standards, <https://ecfr.federalregister.gov/current/title-40/chapter-I/subchapter-C/part-50> (last access: March 2022), 2019.
- US Environmental Protection Agency: Air Data: Air Quality Data Collected at Outdoor Monitors Across the US, https://aqs.epa.gov/aqswb/airdata/download_files.html#Meta/ (last access: May 2021), 2021a.
- US Environmental Protection Agency: Criteria pollutants National Tier 1 for 1970–2020, <https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data> (last access: May 2021), 2021b.
- van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, *Earth Syst. Sci. Data*, 9, 697–720, <https://doi.org/10.5194/essd-9-697-2017>, 2017.
- Vu, T. V., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., and Harrison, R. M.: Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique, *Atmos. Chem. Phys.*, 19, 11303–11314, <https://doi.org/10.5194/acp-19-11303-2019>, 2019.
- Wang, Y. X., McElroy, M. B., Jacob, D. J., and Yantosca, R. M.: A nested grid formulation for chemical transport over Asia: Applications to CO, *J. Geophys. Res.-Atmos.*, 109, D22307, <https://doi.org/10.1029/2004JD005237>, 2004.
- Wells, B., Dolwick, P., Eder, B., Evangelista, M., Foley, K., Mannshardt, E., Misenis, C., and Weishampel, A.: Improved estimation of trends in US ozone concentrations adjusted for inter-annual variability in meteorological conditions, *Atmos. Environ.*, 248, 118234, <https://doi.org/10.1016/j.atmosenv.2021.118234>, 2021.
- Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, *J. Roy. Stat. Soc. B*, 73, 3–36, 2011.
- Xie, Y., Wang, Y., Dong, W., Wright, J. S., Shen, L., and Zhao, Z.: Evaluating the Response of Summertime Surface Sulfate to Hydroclimate Variations in the Continental United States: Role of Meteorological Inputs in the GEOS-Chem Model, *J. Geophys. Res.-Atmos.*, 124, 1662–1679, <https://doi.org/10.1029/2018JD029693>, 2019.
- Xie, Y., Lin, M., Decharme, B., Delire, C., Horowitz, L. W., Lawrence, D. M., Li, F., and Séférian, R.: Tripling of western US particulate pollution from wildfires in a warming climate, *P. Natl. Acad. Sci. USA*, 119, e2111372119, <https://doi.org/10.1073/pnas.2111372119>, 2022.
- Zhai, S., Jacob, D. J., Wang, X., Shen, L., Li, K., Zhang, Y., Gui, K., Zhao, T., and Liao, H.: Fine particulate matter (PM_{2.5}) trends in China, 2013–2018: separating contributions from anthropogenic

- emissions and meteorology, *Atmos. Chem. Phys.*, 19, 11031–11041, <https://doi.org/10.5194/acp-19-11031-2019>, 2019.
- Zhai, S., Jacob, D. J., Wang, X., Liu, Z., Wen, T., Shah, V., Li, K., Moch, J. M., Bates, K. H., Song, S., Shen, L., Zhang, Y., Luo, G., Yu, F., Sun, Y., Wang, L., Qi, M., Tao, J., Gui, K., Xu, H., Zhang, Q., Zhao, T., Wang, Y., Lee, H. C., Choi, H., and Liao, H.: Control of particulate nitrate air pollution in China, *Nat. Geosci.*, 14, 389–395, <https://doi.org/10.1038/s41561-021-00726-z>, 2021.
- Zhang, H., Yuan, H., Liu, X., Yu, J., and Jiao, Y.: Impact of synoptic weather patterns on 24 h-average PM_{2.5} concentrations in the North China Plain during 2013–2017, *Sci. Total Environ.*, 627, 200–210, <https://doi.org/10.1016/j.scitotenv.2018.01.248>, 2018.
- Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., Liu, W., Ding, Y., Lei, Y., Li, J., Wang, Z., Zhang, X., Wang, Y., Cheng, J., Liu, Y., Shi, Q., Yan, L., Geng, G., Hong, C., Li, M., Liu, F., Zheng, B., Cao, J., Ding, A., Gao, J., Fu, Q., Huo, J., Liu, B., Liu, Z., Yang, F., He, K., and Hao, J.: Drivers of improved PM_{2.5} air quality in China from 2013 to 2017, *P. Natl. Acad. Sci. USA*, 116, 24463–24469, <https://doi.org/10.1073/pnas.1907956116>, 2019.
- Zhang, Y., Vu, T. V., Sun, J., He, J., Shen, X., Lin, W., Zhang, X., Zhong, J., Gao, W., Wang, Y., Fu, T. M., Ma, Y., Li, W., and Shi, Z.: Significant Changes in Chemistry of Fine Particles in Wintertime Beijing from 2007 to 2017: Impact of Clean Air Actions, *Environ. Sci. Technol.*, 54, 1344–1352, <https://doi.org/10.1021/acs.est.9b04678>, 2020.
- Zhao, Y., Zhang, K., Xu, X., Shen, H., Zhu, X., Zhang, Y., Hu, Y., and Shen, G.: Substantial Changes in Nitrate Oxide and Ozone after Excluding Meteorological Impacts during the COVID-19 Outbreak in Mainland China, *Environ. Sci. Technol. Lett.*, 7, 402–408, <https://doi.org/10.1021/acs.estlett.0c00304>, 2020.
- Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng, Y., He, K., and Zhang, Q.: Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions, *Atmos. Chem. Phys.*, 18, 14095–14111, <https://doi.org/10.5194/acp-18-14095-2018>, 2018.
- Zhong, Q., Ma, J., Shen, G., Shen, H., Zhu, X., Yun, X., Meng, W., Cheng, H., Liu, J., Li, B., Wang, X., Zeng, E. Y., Guan, D., and Tao, S.: Distinguishing Emission-Associated Ambient Air PM_{2.5} Concentrations and Meteorological Factor-Induced Fluctuations, *Environ. Sci. Technol.*, 52, 10416–10425, <https://doi.org/10.1021/acs.est.8b02685>, 2018.
- Zurbenko, I. G.: Detecting and tracking changes in ozone air quality, *Air and Waste*, 44, 1089–1092, <https://doi.org/10.1080/10473289.1994.10467303>, 1994.