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

Minghao Qiu[1,*], Corwin Zigler[2], and Noelle Selin[1,3]

[1]Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, USA
[2]Departments of Statistics and Data Science and Women's Health, University of Texas, Austin, USA
[3]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, USA
[*]Current address: Department of Earth System Science, Stanford University, USA (mhqiu@stanford.edu)

Correspondence: Minghao Qiu (mhqiu@mit.edu)

1 **Abstract.** Evaluating the influence of anthropogenic emissions changes on air quality requires accounting for the influence of

2 meteorological variability. Statistical methods such as multiple linear regression (MLR) models with basic meteorological vari-

3 ables are often used to remove meteorological variability and estimate trends in measured pollutant concentrations attributable

4 to emissions changes. However, the ability of these widely-used statistical approaches to correct for meteorological variability

5 remains unknown, limiting their usefulness in the real-world policy evaluations. Here, we quantify the performance of MLR

6 and other quantitative methods using simulations from a chemical transport model, GEOS-Chem, as a synthetic dataset. Focus-

7 ing on the impacts of anthropogenic emissions changes in the US (2011 to 2017) and China (2013 to 2017) on $PM_{2.5}$ and O_3,

8 we show that widely-used regression methods do not perform well in correcting for meteorological variability and identifying

9 long-term trends in ambient pollution related to changes in emissions. The estimation errors, characterized as the differences

10 between meteorology-corrected trends and emission-driven trends under constant meteorology scenarios, can be reduced by

11 30%-42% using a random forest model that incorporates both local and regional scale meteorological features. We further

12 design a correction method based on GEOS-Chem simulations with constant emission input and quantify the degree to which

13 anthropogenic emissions and meteorological influences are inseparable, due to their process-based interactions. We conclude

14 by providing recommendations for evaluating the impacts of anthropogenic emissions changes on air quality using statistical

15 approaches.

1 Introduction

Researchers and policy makers 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 diameter less than 2.5 microns ($PM_{2.5}$) have been observed in many countries that adopted policies to limit anthropogenic emissions such as SO_2 and NO_x, including the US (McClure and Jaffe, 2018) and China (Zhang et al., 2019). As information on anthropogenic emissions are often unavailable or very uncertain, researchers and policy makers often rely on the trends in measured air pollutants to assess the effects of polices. Attributing trends in observed concentrations to anthropogenic emissions 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 emissions 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 emissions-driven linear trends in measured concentrations of $PM_{2.5}$ and ozone (O_3), 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 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 three-year average of the $PM_{2.5}$ concentration to determine compliance with air quality standards (European Union, 2020; U.S. 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 O_3 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 emissions 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 O$_3$ from 2013 to 2019 in China were driven by anthropogenic emissions changes. They first use MLR to predict the PM$_{2.5}$ and O$_3$ concentrations with meteorological variables, and then interpret the residuals of the MLR model as signals resulting from emissions 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 pollutant 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 O$_3$ (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 calculates the "weather-normalized" concentration for each day with 1000 sets of meteorological fields drawn from the historical meteorological data. They found that the decrease of PM$_{2.5}$ during 2013 to 2017 was largely driven by emissions reductions, although the magnitude of reduction is smaller when correcting for meteorological variability.

Despite a large number of papers which 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 emissions 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 emissions changes. This is not completely possible, as the impacts of meteorological variability

74 on pollutant concentration will also vary depending on the emissions. The degree to which this interaction affects the ability to

75 calculate emissions-related trends under changing meteorology also remains unknown.

76   Here, we conduct a model experiment to evaluate the performance of widely-used statistical models in correcting for me-

77 teorological variability and estimating emissions-driven trends in air quality (see figure 1). We focus on the impacts of an-

78 thropogenic emissions changes on annual $PM_{2.5}$ and summer O_3 in the US (2011-2017) and China (2013-2017), two periods

79 well-studied in previous literature. Using a 3-D atmospheric chemical transport model GEOS-Chem, we simulate two sets of

80 scenarios – "observational scenarios" with assimilated meteorological inputs (with interannual variability) and "counterfactual

81 scenarios" with constant meteorological inputs. Using simulated daily concentrations in the observational scenarios, we es-

82 timate meteorology-corrected trends for each grid cell from regression models using different statistical correction methods.

83 We then compare the derived trends with the emissions-driven trends in the counterfactual scenarios (which are free of mete-

84 orological variability by design), calculating the resulting "error" in trend estimation. We further design a correction method

85 based on GEOS-Chem constant emission simulations, and use it to quantify the degree to which attribution to meteorology and

86 emissions separately is possible. Finally, we apply the different statistical correction methods to observational data from sur-

87 face monitoring networks in the US and China, discussing the variability across different methods. We conclude by providing

88 recommendations for techniques to evaluate air pollution policies under changing meteorological conditions.

89 ## 2   Method

90 ### 2.1   GEOS-Chem

91 GEOS-Chem is a global three-dimensional chemical transport model driven by assimilated meteorological data from the God-

92 dard Earth Observation System (GEOS-5) of the NASA Global Modeling and Assimilation Office (GMAO) (Bey et al. (2001),

93 http://www.geos-chem.org/). The simulation of $PM_{2.5}$ in GEOS-Chem represents an external mixture of secondary inorganic

94 aerosols, carbonaceous aerosols, sea salt, and dust aerosols. GEOS-Chem includes detailed O_3-NO_x-volatile organic carbon

95 (VOC)-aerosol-halogen tropospheric chemistry (Travis et al., 2016; Sherwen et al., 2016). The GEOS-Chem model has been

96 previously used to study the changes in $PM_{2.5}$ and O_3 during our studied periods, and model simulations have been shown to

97 be consistent with the observed concentrations (e.g., see Li et al. (2017a); Xie et al. (2019) for the US, and Li et al. (2018);

98 Lu et al. (2019); Zhai et al. (2021) for China). Studies in both regions show that the GEOS-Chem model is able to reproduce

99 the spatial, seasonal, and interannual variability and the long-term trends in observed pollutant concentrations, despite biases

100 in absolute concentrations in certain species and regions (Heald et al., 2012; Travis et al., 2016; Tian et al., 2021).

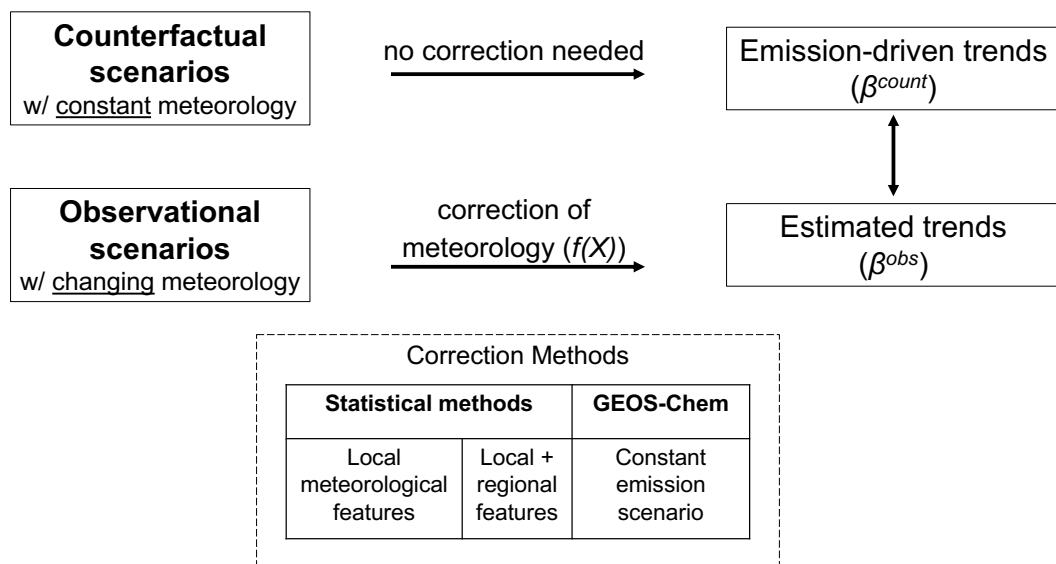



Figure 1. Overview of research methodology. Terms and coefficients are linked to the associated terms in equation 1 and table 1.

101   We use GEOS-Chem version 12.3.0 with a horizontal resolution of $0.5° \times 0.625°$ in North America and Asia (Wang et al.,

102   2004). For each scenario, we first conduct a global run at a horizontal resolution of $4° \times 5°$, with a 12 month spin-up. These

103   global runs are then used as the boundary conditions for nested simulations in US and Asia with finer resolution of $0.5° \times 0.625°$.

104   **2.2   GEOS-Chem scenarios**

105   Table 1 shows the simulations included in our model experiments. We simulate two sets of scenarios – "observational scenarios"

106   with interannual variability in meteorology and "counterfactual scenarios" with constant meteorological inputs. Both scenarios

107   use the same emissions inventory as input (see section 2.3). For each grid cell, we estimate the linear trends in pollutant

108   concentrations from simulated daily $PM_{2.5}$ and O_3 concentrations. We focus on the daily 24-hour average $PM_{2.5}$ over all

109   seasons, and the maximum daily average 8-hour (MDA8) O_3 in summer (June, July, August). Our focus on the three summer

110   months is consistent with many previous studies (e.g., Shen et al. (2015)), although this may not capture the peak ozone

111   season for certain regions of the US and China. Our GEOS-Chem simulations use meteorological fields from the Modern-

112   Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro et al., 2017). We aggregate the

113   hourly meteorological data for consistency with the pollutant concentrations: a 24-hour average for $PM_{2.5}$ analysis and the

114   corresponding 8-hour average for O_3. Meteorological features that are used in the statistical models can be found in section

115   2.4.

2.2.1 Observational scenarios

Observational scenarios simulate $PM_{2.5}$ 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 figure A1 and A2, the simulated concentrations in $PM_{2.5}$ and O_3 largely reproduce the daily variability in observed pollutant concentrations. The linear trends in simulated $PM_{2.5}$ and O_3 concentrations in the observational scenario are largely consistent with trends of the measured concentrations. For example, the average trend ($\pm$one standard deviation) in the US is -0.27$\pm$0.30 μg^{-3}/year (observation) and -0.39$\pm$0.24 ppb/year (GEOS-Chem) for $PM_{2.5}$, and -0.91$\pm$0.98 ppb/year (observation) and -1.02$\pm$0.83 ppb/year (GEOS-Chem) for O_3. The only exception is that our model cannot reproduce the increasing $PM_{2.5}$ trends in the Northwest US because we do not consider interannual variability in biomass burning emissions.

2.2.2 Counterfactual scenarios

Counterfactual scenarios simulate $PM_{2.5}$ 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 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 emissions changes alone. In a sensitivity analysis, we also simulate the counterfactual scenario for China using the meteorological fields at the end year 2017 (at 4$\times$5 degree resolution, due to computational constraints). We find that the linear trend in $PM_{2.5}$ and O_3 for each grid cell is highly consistent in the counterfactual scenarios across the choice of the meteorological years (see figure A5).

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 much 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.

GEOS-Chem scenarios	Emissions inventory	Meteorological fields	Trend estimates	Meteorological correction
Counterfactual scenarios	**Changing** 2011-2017 (US) 2013-2017 (China)	**Constant** 2011 (US) 2013 (China)	β^{count}	
Observational scenarios	**Changing** 2011-2017 (US) 2013-2017 (China)	**Changing** 2011-2017 (US) 2013-2017 (China)	$\beta^{uncorrected}$	No correction
			β^{MLR}	Linear combination of local features
			β^{GAM}	GAM using local features
			β^{RF}	RF using local features
			$\beta^{LASSO-regional}$	LASSO using local and regional features
			$\beta^{RF-regional}$	RF using local and regional features
			β^{gc}	Use simulations from constant emissions scenarios
Constant emissions scenarios	**Constant** 2011 (US) 2013 (China)	**Changing** 2011-2017 (US) 2013-2017 (China)		

Table 1: Overview of GEOS-Chem scenarios and meteorological correction methods.

2.3 Emissions inventory

For the US, we use the National Emissions Inventory 2011 (NEI 2011) as a baseline emissions inventory and scale the emissions in 2012 to 2017 to match the annual total emissions each year (U.S. Environmental Protection Agency, 2021b). For China, we use the monthly Multi-resolution Emission Inventory for China (MEIC) during 2013 to 2017 (Li et al., 2017b; Zheng et al., 2018). During the studied time periods, the US and China experienced dramatic decreases in anthropogenic emissions, particularly in SO_2 and NO_x. In the US, total anthropogenic emissions of SO_2 decreased by 57% and NO_x emissions decreased by 26% during 2011 to 2017 (see figure A3). In China, anthropogenic SO_2 emissions decreased by 59% and NO_x emissions decreased by 21% during the 2013-2017 period (see figure A4).

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 O_3 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 Global Fire Emissions Database (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:

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

where y_{it} denotes the PM$_{2.5}$ or O$_3$ concentration at grid cell i on day t. t is the time index (e.g., in the US, $t=1$ for January 1st, 2011 and $t=2$ for January 2nd, 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 of pollutant concentrations that are not related to the meteorological variability (e.g., seasonal cycle in O$_3$ 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 O$_3$ concentration for grid cell i estimated with the standard ordinary least square method. We use the absolute differences $|\beta_i^{obs} - \beta_i^{count}|$ 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 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 manuscript, and the performance of the other methods can be found in table A1 and A2. Note that the time fixed effects are modelled differently in RF models due to the estimation procedure. More details on the implementation of RF can be found in SI.

We use the following ten 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 level). 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, top soil 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 O$_3$ concentrations in the GEOS-Chem model and may contain information that help explain 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 synoptic scale (i.e. large scale weather systems that span over 1000 kilometers 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

197 show good prediction performances with high dimensional data inputs. We estimate the following equations:

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

199 where $g_i()$ denotes the functional form fitted by LASSO or RF. X_{it} again denotes the local meteorology features for grid cell

200 i on day t. Z_t denotes the regional scale meteorology features including the meteorological features for every grid cell in the

201 US on day t (98 cells in 4×5 degrees; we choose a relatively coarse resolution due to computational cost). Meteorological

202 information in each location in the US may help explain the pollutant concentrations in grid cell i. In total, we have 10 local

203 features (X_{it}) and 10×98=980 regional scale features (Z_t). The coefficient β_i^{obs} is obtained with the double machine learning

204 approach by Chernozhukov et al. (2018). In particular, the hyper-parameters and coefficients of LASSO and RF are selected

205 and fitted using 4-fold cross-validation to avoid the "overfitting risk". More details on the implementation of LASSO and RF

206 can be found in SI.

207 2.5 Correction approach using GEOS-Chem constant emissions scenario

208 We further design and evaluate an approach to correct for meteorology variability with GEOS-Chem simulations (referred

209 to as "constant-emis" approach). The "constant-emis" approach uses GEOS-Chem simulations with constant anthropogenic

210 emissions and changing meteorological fields ("constant emissions scenarios" in table 1). All years in the constant emissions

211 scenario use anthropogenic emissions of the start year (2011 for US, 2013 for China). We estimate the following equations:

$$y_{it} = \beta_i^{gc} \times t + SIM_{it} + \eta_{it} + \epsilon_{it} \tag{3}$$

213 where SIM_{it} denotes the simulated concentrations on day t in grid cell i in the constant emissions scenarios. SIM_{it} serves a

214 similar purpose as the term "$f_i(X_{it})$" in equation 1, but comes from the GEOS-Chem simulation. Some previous studies have

215 also used model simulations with constant emissions input as a way to characterize meteorological variability (Zhong et al.,

216 2018; Zhao et al., 2020). β_i^{gc} is the estimated meteorology-corrected trend in PM$_{2.5}$ or O$_3$ concentration using this model-based

217 correction method.

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

219 rological variability as simulated in GEOS-Chem (as SIM_{it} are directly taken from GEOS-Chem). Therefore, the difference

220 between the trend estimates (β^{gc}) and counterfactual trends (β^{count}) provides a conceptual minimum for estimation errors us-

221 ing the framework of equation 1 to perform meteorological corrections. The commonly-used framework of equation 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 emissions inventory). We use $|\beta_i^{gc} - \beta_i^{count}|$ 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 Systems administered by the US EPA (U.S. Environmental Protection Agency, 2021c). We use the daily 24-hour average of $PM_{2.5}$ concentrations for all months and the daily maximum 8-hour average (MDA8) O_3 concentrations for June, July and August. Figure A1 shows the locations, trends in measured concentrations, and correlations between GEOS-Chem simulations and measured concentrations.

The surface air quality measurements in China come from the monitoring network from 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-hour average of $PM_{2.5}$ concentrations and the MDA8 O_3 concentrations for summer. Figure A2 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 analysis 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 2C show the trends in $PM_{2.5}$ and O_3 concentrations in the counterfactual scenarios in the US. When holding meteorological fields constant across years, decreasing trends in the simulated $PM_{2.5}$ concentrations across the US result from decreasing anthropogenic emissions. The counterfactual scenario also shows negative linear trends in O_3 concentrations in all but three grid cells in the West. Increases in summer O_3 in these locations result from the non-linear relationship between O_3 concentrations and NO_x emissions.

248   Figure 2B and 2D show the degree to which different meteorological correction methods can recover the emissions-driven

249   trends in the counterfactual scenarios. When no correction for meteorology is performed ("uncorrected" in figure 2B), we

250   observe large estimation errors in trend estimates over the Northeast and Southern US by up to 0.25 μg m^{-3}/year, an error

251   that is 50% of the counterfactual trends. We find that the widely-used MLR method does not help reduce these errors in PM$_{2.5}$

252   trend attributions. MLR has a modest impact on reducing the errors in Northeast US, but it does not decrease the errors over

253   the Southern US and leads to even higher errors over Midwest. Nationwide, the average magnitude of errors (relative to the

254   counterfactual scenario) *increases* with the MLR correction (0.083 μg m^{-3}/year) compared to the uncorrected case (0.066 μg

255   m^{-3}/year). Among the five methods, we find that the RF model using both local and regional scale features ("RF-regional" in

256   figure 2) offers the best performance in recovering the trends in the counterfactual scenarios and is the only method that yields

257   smaller errors than the uncorrected case (the nationwide average error decreased by 0.019 μg m^{-3}/year, or 28% less). The

258   RF-regional model also outperforms the RF-local and LASSO-regional models, suggesting the importance of considering non-

259   linearity, interactions between different meteorological features, and regional meteorology information in correctly adjusting

260   for the impacts of meteorology.

261   Meteorological variability also has a substantial influence on the summertime O$_3$ trends in the US during this period (as

262   shown in figure 2D). Relative to the counterfactual scenario, the uncorrected O$_3$ trends are biased by over 1-2 ppb/year in large

263   areas of California, Midwest and Southern US (as much as 320% of the counterfactual trends). This is largely driven by the fact

264   that the 2011 and 2012 summer was particularly hot in these regions and led to higher concentrations of O$_3$ at the beginning of

265   this 7-year period (see figure A7 for the Southern and Midwest US). Therefore, failure to correct for meteorological variability

266   results in much more negative trend estimates in the O$_3$ concentrations in these areas compared to the counterfactual scenario

267   (see figure A6). Meteorology corrections with MLR or GAM help reduce these estimation errors substantially (nationwide

268   average error is reduced by 51% using MLR or 57% using GAM compared to uncorrected trends), while large errors still

269   persist in the Midwest and South. Similar to the case of PM$_{2.5}$, the RF-regional model offers the best performance in correcting

270   for meteorological variability (the national average error is further reduced by 42%, compared to MLR), and it is especially

271   helpful in reducing the errors over the Midwest and South (regional average error is reduced by 64% and 44%, respectively,

272   compared to MLR).

273   **3.2   Performance of different correction methods: China (2013-2017)**

274   Figure 3A and 3C show the trends in PM$_{2.5}$ and O$_3$ concentrations in the counterfactual scenarios in China. We find a sub-

275   stantial decline in simulated PM$_{2.5}$ concentration during 2013 to 2017, particularly in eastern and central China. In contrast,

276   there is little change in the simulated PM$_{2.5}$ concentrations in western China in the counterfactual scenario, where PM$_{2.5}$ is

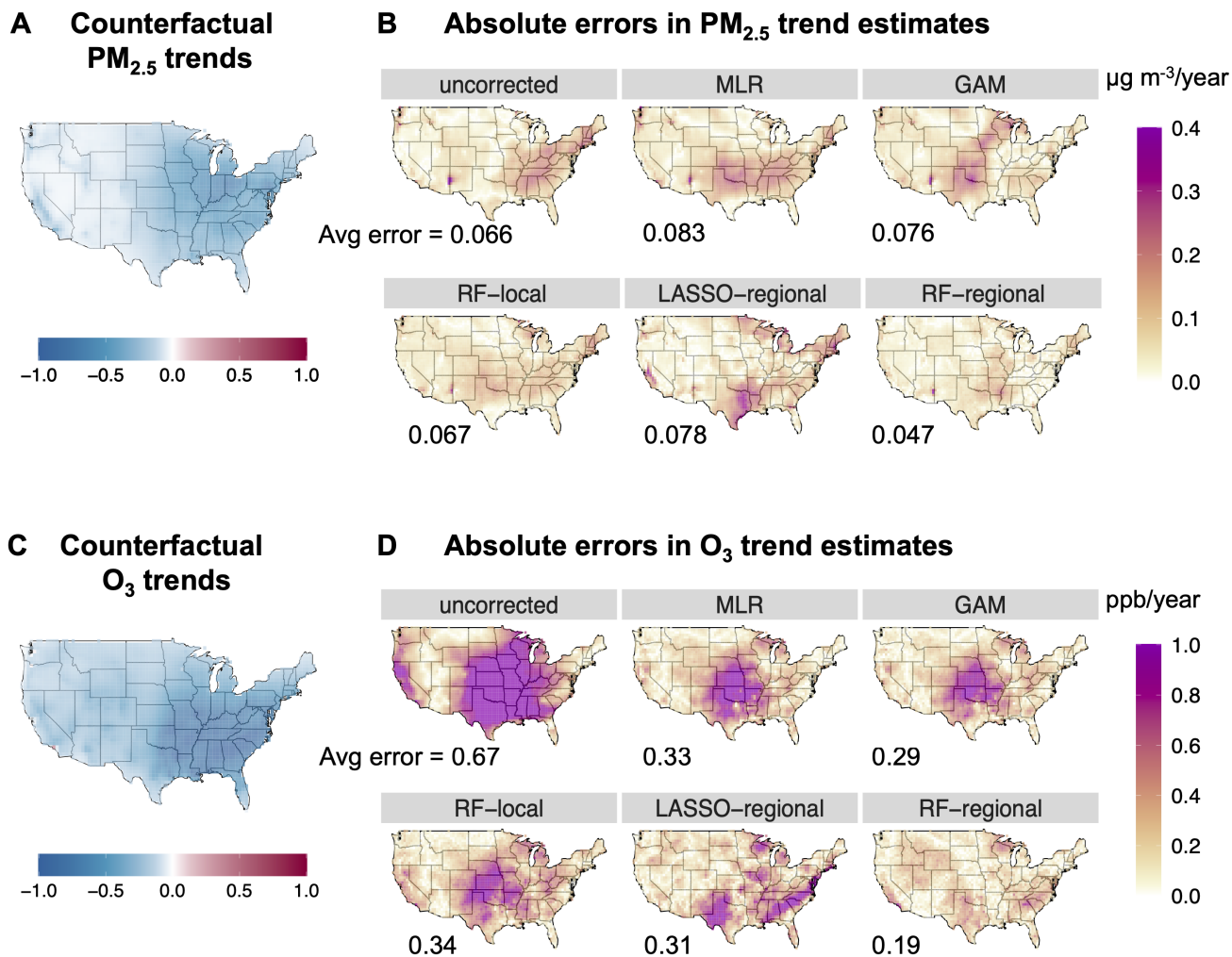



Figure 2. Trend estimates of daily annual PM$_{2.5}$ (Panels A and B) and summer O$_3$ (C and D) in the US. Panels A and C show trend estimates under the counterfactual scenario (β^{count}). Panels B and D show the absolute magnitude of errors of trend estimates under different correction methods compared with the counterfactual scenarios ($|\beta^{obs} - \beta^{count}|$). The average of the absolute errors for each method is shown in the figure. Unit of trend estimate is μg m^{-3}/year for PM$_{2.5}$ or ppb/year for O$_3$.

277 dominated by dust species largely driven by natural processes (see figure A9). For summer O_3, there are decreasing trends in

278 the counterfactual scenario in most parts of China, except for North China and some urban areas. This is largely consistent with

279 previous studies that attempt to attribute emissions-related changes in O_3 concentrations during this period based on modeling

280 or observational data (Li et al., 2018, 2020; Lu et al., 2020).

281   Figure 3B shows the magnitude of estimation errors in the trend estimates of annual $PM_{2.5}$ in China under different correction

282 methods. We find the underlying meteorological variability has a substantial impact on $PM_{2.5}$ trends in China during this

283 period. We observe large differences between the uncorrected and counterfactual trends in simulated $PM_{2.5}$ concentrations,

284 particularly in Northwest and Northeast China. Similar to the model experiments in the US, we find that MLR and GAM

285 methods fail to correct for this underlying meteorological variability and lead to further increases in estimation errors in many

286 locations. Relative to the counterfactual scenario, the nationwide average error increases to 0.90 μg m^{-3}/year with MLR

287 and 1.06 μg m^{-3}/year with GAM (compared to 0.89 μg m^{-3}/year with no correction). We find that the RF-regional model

288 recovers the counterfactual trends better than other methods (nationwide average error: 0.64 μg m^{-3}/year; an improvement by

289 30% relative to MLR), but it is still not able to correct for the persistent estimation errors over Northwest China. We further

290 analyze the performance of correction methods for the different component species of $PM_{2.5}$. As shown in figure A10 and

291 A11, the MLR model is particularly unable to correct for the impacts of meteorological variability on nitrate and dust species.

292 Compared with MLR, the RF-regional model better corrects for the impacts of meteorology on secondary organic aerosol

293 species in South and Central China and ammonium in Northeast, but only yields modest improvement in correcting for the

294 errors in dust concentrations over Northwest China (see figure A12). In a sensitivity analysis, we use an approach that first fits

295 RF-regional models of each individual $PM_{2.5}$ species, and then combines predictions for each species to derive trend estimates.

296 The results are largely similar to the main approach that directly fits the total $PM_{2.5}$ concentration (see figure A13).

297   Figure 3D shows the magnitude of errors in the trend estimates for summer O_3 under different correction methods in

298 China. We find that the MLR model only modestly reduces the estimation errors compared to the uncorrected cases, and the

299 RF-regional model offers the best overall performance. The nationwide average error is reduced to 0.28 ppb/year using the RF-

300 regional model (relative to 0.43 ppb/year uncorrected and 0.41 ppb/year with MLR). Similar to the evaluation of summer time

301 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

302 model. Surprisingly, the LASSO-regional model performs the worst in recovering the counterfactual trends. Compared to the

303 US case, we find that the impacts of meteorological variability on O_3 and the method performances are much more spatially

304 heterogeneous (see figure A6, A8), which may be partially due to the more heterogeneous O_3 regimes in China during this

305 period.

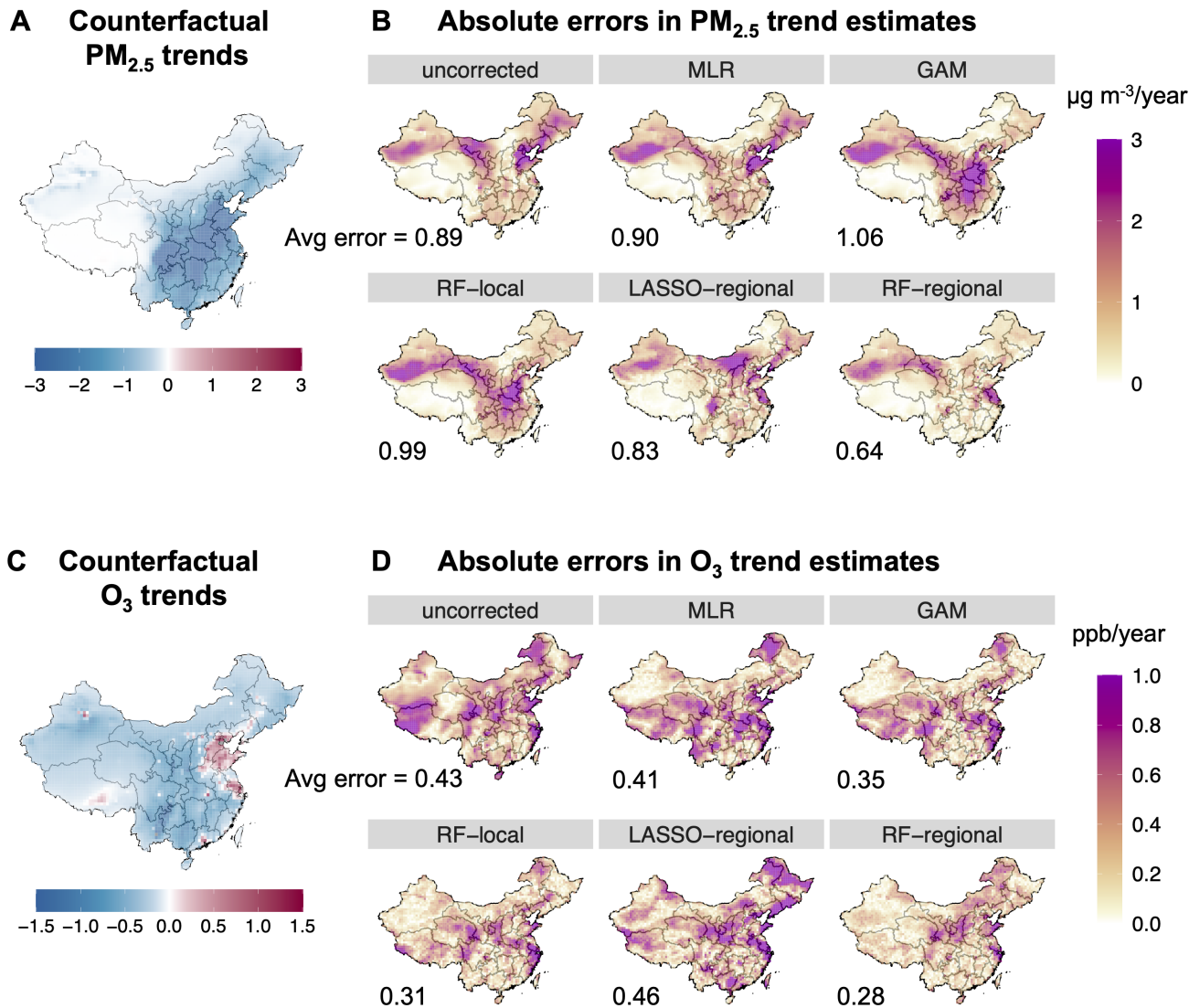



Figure 3. Trend estimates of daily annual PM$_{2.5}$ (Panels A and B) and summer O$_3$ (C and D) in China. Panels A and C show trend estimates under the counterfactual scenario (β^{count}). Panels B and D show the absolute magnitude of errors of trend estimates under different correction methods compared with the counterfactual scenarios ($|\beta^{obs} - \beta^{count}|$). The average of the absolute errors for each method is shown in the figure. The unit of the trend estimate is μg m^{-3}/year for PM$_{2.5}$ or ppb/year for O$_3$.

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

In our model experiments in both 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 not only on 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 of 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 approach. Figure 4 shows the estimation errors of trend estimates using the constant emissions scenarios simulated by GEOS-Chem. We focus on the trends in summer O_3 in the US and annual $PM_{2.5}$ 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 figure 4A and 4C), trends evaluated using the constant-emis approach are very similar to the trends in the counterfactual scenarios. The national average error of trend estimates is only 0.04 ppb/year for the O_3 trends in the US (relative to 0.33 ppb/year under MLR or 0.19 ppb/year under RF-regional), and only 0.08 μg m^{-3}/year for the $PM_{2.5}$ trends in China (relative to 0.91 μg m^{-3}/year under MLR or 0.64 μg m^{-3}/year under RF-regional).

However, the estimation errors calculated above are still non-negligible and can be large in certain regions. As shown in Figure 4B and 4D, 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 $PM_{2.5}$ in China estimated from the observaional 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

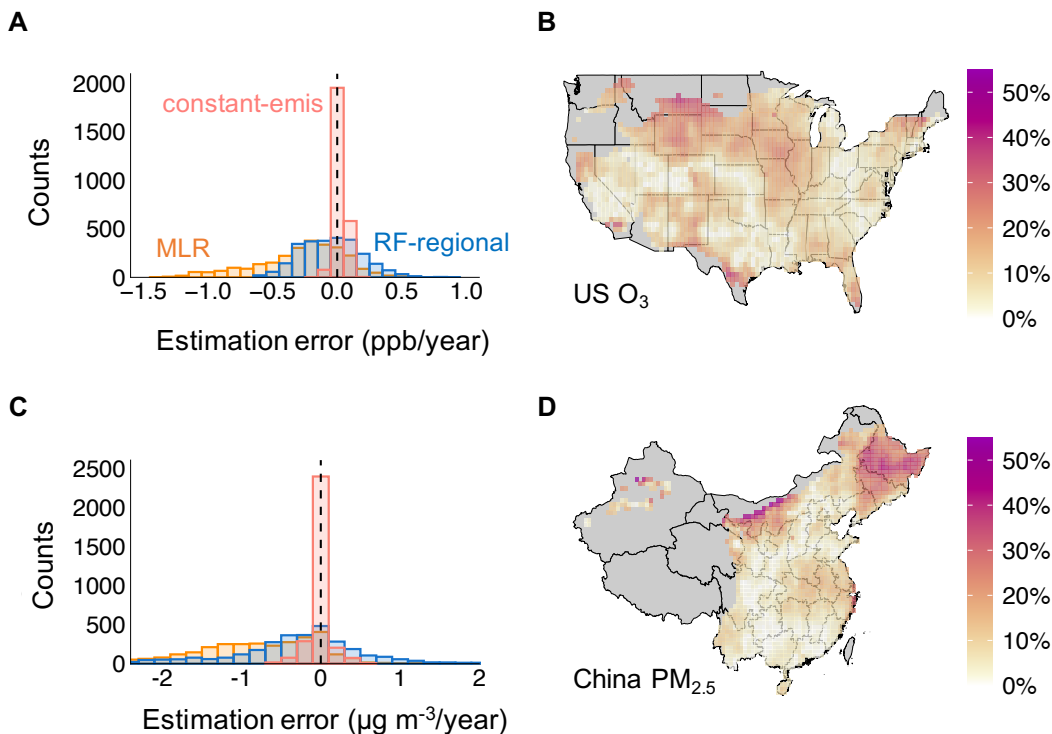



Figure 4. Panels A and C show the histogram of estimation errors in trend estimates assessed using MLR, RF-regional and constant-emis. Panels B and D show the errors assessed with the constant-emis method as a percentage of the trends in the counterfactual scenario ($|\beta^{gc} - \beta^{count}|/|\beta^{count}|$). Panels B and D only show grid cells with a trend in the counterfactual scenarios >0.2 ppb/year or >0.2 μg m^{-3}/year; remaining grid cells are shown in gray. Panels A and B illustrate the summer O$_3$ trends in the US. Panels C and D illustrate the annual PM$_{2.5}$ trends in China.

334 different meteorological correction methods to the observational data, our analysis reveals that the choice of methods for

335 meteorological correction can yield very different results for certain regions. For example, the regional average uncorrected O$_3$

336 trend is -1.49 ppb/year and -1.15 ppb/year in Midwest and Southern US, respectively, which overestimates the reductions in

337 concentrations attributable to anthropogenic emissions changes (figure 5A). Correcting for the meteorological variability with

338 MLR model yields regional average trend at -0.54 ppb/year in Midwest (a decrease by 53% in magnitude relative to uncorrected

339 trends) and -0.71 ppb/year in the Southern US (a decrease by 52%). RF-regional model further reduces the absolute magnitude

340 of the declines in O$_3$ attributable to emissions reductions to -0.02 ppb/year for Midwest and -0.40 ppb/year for the Southern

341 US. Importantly, these patterns are consistent with the results from our model experiments in these regions: the RF-regional

342 model also estimates a much less negative emissions-driven trend in the Southern US compared to the uncorrected case and

343 MLR estimates in the GEOS-Chem simulations. For the GEOS-Chem simulations, RF-regional estimates are 39% smaller than

344 MLR estimates, and this is comparable to the magnitude changes for the observational data (RF-regional estimates are 44%

345 smaller than MLR). As the RF-regional model outperforms the other correction methods in recovering counterfactual trends in

346 the GEOS-Chem simulations, this potentially also suggests a better performance of RF-regional in recovering the underlying

347 emission-driven trends when applying to the observational data.

348   We find similar consistency of the method performances between observational data and GEOS-Chem simulations in China

349 as well (figure 5B). When applying to the observational data from the surface monitoring network, much smaller reduction

350 of $PM_{2.5}$ concentrations is attributed to anthropogenic emissions changes in the North, Northeast and East of China using

351 the RF-regional model, relative to the MLR estimates. For example, the average emissions-driven trend estimated from the

352 observational data is -4.9 μg m^{-3}/year in Beijing under the RF-regional model, compared with -9.6 μg m^{-3}/year under the

353 MLR model. These patterns are consistent with the patterns of the trend estimates estimated from our GEOS-Chem simulations

354 with different statistical methods.

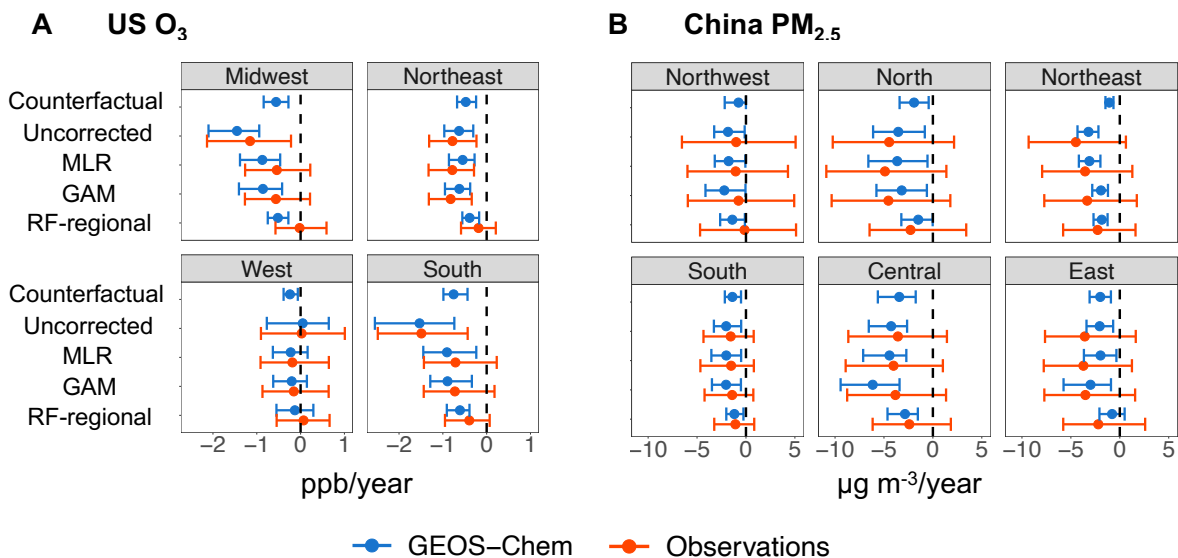



Figure 5. Trends in O_3 in the US (panel A) and $PM_{2.5}$ in China (panel B) estimated from the observational data (red) and GEOS-Chem simulations (blue) under different correction methods. Trends in pollutant concentrations are estimated at the monitor level (for the observational data) or at the grid 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 O_3 trends in the US (unit: ppb/year). Panel B illustrates the annual $PM_{2.5}$ trends in China (unit: μg/m^3/year). 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 U.S. Environmental Protection Agency (2021c) and China's Ministry of Ecology and Environment (2021).

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 emissions changes. Based on our evaluations of either $PM_{2.5}$ or O_3 trends across 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 emissions-driven trends. We propose a random forest model that uses both local and regional meteorological features, which offers the best overall performance in recovering the emissions-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 emissions changes on the decline of 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 figures A14 and A15). 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 emissions- 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 of 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 (U.S. 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 (such as RF-regional) performs notably worse in correcting for the variability in dust-related $PM_{2.5}$ (see figure A12), likely because dust $PM_{2.5}$ is extremely variable, with zero concentration in 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 $PM_{2.5}$ 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 of 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 of the constant-emis approach. Our work only evaluates the statistical and machine learning models in expressions 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 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 statisti-

414 cal models with longer simulations and alternative trend estimates (such as the Theil-Sen estimator). A deeper investigation of

415 the estimation error due to assuming perfect separation between meteorology and emission is also essential for understanding

416 how we should interpret studies that use these statistical methods. For example, further work could explore how these errors

417 will vary by the magnitude of emissions reductions and the chemistry regimes.

5 Recommendations for attributing trends to emissions changes

419 Using statistical methods to causally infer relationships between simulated air pollutant concentrations and anthropogenic

420 emissions is challenging, and doing so in contexts of observational data is even more challenging. Understanding the uncer-

421 tainty of statistical models in characterizing the meteorology-pollution relationship is essential to evaluating the effectiveness

422 of policy interventions with observational data. Here, we make several recommendations to researchers and policy makers

423 based on our analysis.

424 For those who aim to infer causal effects of emissions changes on air quality based on observational data on concentrations

425 and meteorology, we recommend using multiple statistical methods to correct for meteorological variability when evaluating

426 the impacts of policies or interventions on air quality. From our two case studies, we find a relatively large variation between the

427 trend parameters estimated by different statistical methods (especially at the grid cell or monitor level). Some methods perform

428 better in certain locations but not in others (though RF-regional is the best-performing method overall). Using multiple ap-

429 proaches (linear/non-linear and at local/regional scale) may help to quantify uncertainty related to meteorological corrections.

430 These findings also suggest that empirical analyses may benefit from considering the impacts of meteorological variability on

431 air quality separately for each region or even for each monitor location (if data permits), instead of attempting to determine a

432 general relationship between meteorological variability and air pollution over a large spatial domain. Finally, analysts should

433 be particularly cautious when using statistical methods to estimate impacts of anthropogenic emissions on air quality in regions

434 where pollution variability is dominated by meteorologically-influenced environmental processes such as dust emissions, as

435 we consistently show that typical statistical methods (in combination with the standard set of meteorological variables) do not

436 work well in those regions.

437 Due to the non-negligible estimation errors in recovering the counterfactual trends even with the best-performing statistical

438 approach we test, we believe these statistical analyses are most useful in understanding the patterns of anthropogenic emissions

439 on air quality when aggregated across larger spatial areas, rather than providing specific trends for individual monitor locations.

440 There is a higher degree of consistency among the trend estimates across different methods when aggregated at regional level,

441 but assessment at local level is more sensitive to method choice. The absolute magnitude of monitor-level trends needs to be

442 interpreted with caution, considering both the uncertainty from the statistical methods and also the limit of meteorological

443 correction due to ignoring the interactions between meteorology and emissions.

444   Because measured pollutant concentrations are subject to the influence of underlying meteorological variability, many efforts

445 have attempted to correct for the impacts of meteorological variability and use "meteorology-corrected" concentrations and

446 trends to assist in evaluating the effectiveness of air quality policies. Our study evaluates existing methods that aim to correct for

447 the meteorological variability and finds many of these methods do not perform well. This raises potential concerns about the use

448 of "meteorology-corrected" concentrations as targets for policy evaluation. Meteorology-corrected concentrations and trends

449 remain useful metrics to quantify the influence of emissions. However, a more comprehensive evaluation of the effectiveness

450 of policy requires interpreting measurements with all available tools, ideally including both statistical analyses and physical

451 models.

Code and data availability. The GEOS-Chem simulation of different scenarios and the R code for implementing the different statistical methods are available at the following repository.

Author contributions. M.Q. and N.E.S. designed the research. M.Q. performed the statistical analysis and GEOS-Chem modeling simulations. All authors interpreted the results and wrote the paper.

Competing interests. The authors declare no competing interests.

Acknowledgements. We thank Colette Heald and Valerie Karplus for helpful comments and discussions. We thank Yixuan Zheng for assistance with the MEIC emissions inventory. We thank Ke Li for sharing code of step-wise MLR analysis. This publication was supported by US EPA grant RD-835872-01. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication.

461 **Appendix: Supplementary methods**

462 **Implementation of LASSO and RF**

463 As the incorporation of both local and regional features can quickly expand the dimensionality of the feature space, we use the

464 Least Absolute Shrinkage and Selection Operator (LASSO) and the Random Forest (RF) model to assess the importance of

465 regional meteorological features. Both methods are commonly-used approaches with good prediction performance with high

466 dimensional data inputs, and are thus appropriate for an analysis with a large number of regional meteorological features. For

467 these two methods, we rewrite equation 1 as the following:

468 $$y_{it} = \beta_i^{obs} \times t + g_i(X_{it}, Z_t, W_t) + \epsilon_{it} \tag{1}$$

469 where $g_i()$ denotes the functional form fitted by LASSO or RF. X_{it} again denotes the local meteorological features for grid cell

470 i on day t. Z_t denotes the regional scale meteorology features including those for all grid cells in the US on day t (98 cells in

471 4×5 degrees; we choose a relatively coarse resolution due to computational cost). Meteorological information in each location

472 in the US may help explain the pollutant concentrations in grid cell i. In total, we have 10 local features (X_{it}) and 10×98=980

473 regional scale features (Z_t). W_t denotes the day and month variable to model the daily and monthly variability in pollutant

474 concentrations that are unrelated to meteorological variability. For LASSO, we use month-of-year×day-of-month fixed effect

475 (same as all the other methods except for RF), and these fixed effects are not penalized in the LASSO regression. For RF, we

476 use the month-of-year variable (from 1 to 12), and day-of-month variable (from 1 to 31), due to the inefficient performance of

477 RF working with a large number of fixed effects. Thus, the difference between RF and the other methods may also come from

478 the different choices in modeling monthly and daily variability.

479   The coefficient β_i^{obs} is obtained with the following procedure using the double machine learning approach by Chernozhukov

480 et al. (2018):

481   (1) We first partition the time series of $\{y_{it}, X_{it}, Z_t, W_t\}$ into 4 folds. We use 75% of the data as training data and the

482 remaining 25% for predictions. We train the following two models on the training data:

483 $y_{it} = f(X_{it}, Z_t, W_t)$

484   $t = g(X_{it}, Z_t, W_t)$

485 (2) We then apply models $f(.)$ and $g(.)$ to the prediction set to get predictions of y_{it} and t for the rest 25% of the data. The

486 above process is repeated four times to derive predictions for the entire time series (predictions denoted as $\widehat{y_{it}}$ and $\widehat{t}$).

487 (3) We calculate the residuals of each model $\widetilde{y_{it}} = y_{it} - \widehat{y_{it}}$ and $\widetilde{t} = t - \widehat{t}$. The coefficient of interest β_i^{obs} is then calculated

488 as:

489 $$\beta_i^{obs} = \frac{\sum_t \widetilde{t}\widetilde{y_{it}}}{\sum_t \widetilde{t}\widetilde{t}}$$

490 This is equivalent to setting up a linear regression of $\widetilde{y_{it}} \sim \widetilde{t}$ and obtaining the slope coefficients (as shown by Chernozhukov

491 et al. (2018)).

492 The hyper-parameters of RF and LASSO are tuned with 4-fold cross validation. We also perform two sensitivity analyses: 1)

493 with a different spatial resolution for the regional scale features (2×2.5 degrees instead of 4×5 degrees), and 2) with different

494 numbers of folds to estimate the trend coefficients. Our results are similar across these sensitivity analyses (see figure A16).

495 The double machine learning framework involves a sample partition procedure (steps (1) and (2) above). This procedure,

496 however, does not fit the purpose of including time fixed effects in the LASSO model (as randomly partitioned training and

497 test sets could have very unbalanced number of observations from a given month-day pair). Therefore, step (1) and (2) are only

498 implemented for the RF model, and coefficients of the LASSO model are directly derived from step (3) without sample splitting.

499 This is acceptable for the LASSO model as the risk of "overfitting" has already been eliminated by using the tuned penalizing

500 factor (i.e. the hyper-parameters) derived from a 4-fold cross-validation. It is important to note that we quantify the performance

501 of RF and other methods using the differences between "meteorology-corrected" trends (β^{obs}) and the counterfactual trends

502 (β^{count}), instead of their performance in predicting the pollutant concentration. Therefore, if the RF model "overfits the data",

503 it would actually result in a large error, because the overly fit RF model would attribute all variability of $PM_{2.5}$ and O_3 to the

504 meteorological variables and estimate a close-to-zero trend.

Model	Annual PM$_{2.5}$ in the US			Summer O$_3$ in the US		
	average error	median relative error	cells with relative error >50%	average error	median relative error	cells with relative error >50%
No correction	0.066	28%	27%	0.67	154%	84%
MLR (5 features)	0.092	43%	44%	0.38	84%	71%
MLR (10 features)	0.083	40%	40%	0.33	71%	64%
Quadratic	0.088	40%	42%	0.29	60%	58%
Cubic	0.075	39%	41%	0.28	60%	58%
Spline	0.076	40%	41%	0.28	61%	59%
GAM	0.076	40%	43%	0.29	61%	58%
RF-local	0.067	33%	39%	0.34	78%	70%
LASSO-regional	0.078	31%	33%	0.31	68%	65%
RF-regional	0.047	25%	23%	0.19	46%	47%

Table A1. Estimation errors of trend estimates in the US under different correction methods. The average estimation errors, median relative error, and fraction of grid cells with relative error greater than 50% are shown in the table. Relative errors are calculated as the ratio of estimation error to the trend estimate in the counterfactual scenario. MLR (5 features) only use temperature, precipitation, humidity, and surface wind speed (U,V directions) as the meteorological features.

Model	Annual PM$_{2.5}$ in China			Summer O$_3$ in China		
	average error	median relative error	cells with relative error >50%	average error	median relative error	cells with relative error >50%
No correction	0.89	224%	77%	0.43	95%	74%
MLR (5 features)	1.07	193%	80%	0.42	90%	68%
MLR (10 features)	0.90	159%	79%	0.41	85%	68%
Quadratic	1.00	142%	82%	0.36	76%	62%
Cubic	1.07	143%	82%	0.34	68%	59%
Spline	1.08	140%	84%	0.33	69%	59%
GAM	1.06	139%	82%	0.35	72%	59%
RF-local	0.99	172%	82%	0.31	64%	58%
LASSO-regional	0.83	184%	75%	0.46	98%	73%
RF-regional	0.64	152%	67%	0.28	61%	58%

Table A2. Estimation errors of trend estimates in China under different correction methods. The average estimation errors, median relative error, and fraction of grid cells with relative error greater than 50% are shown in the table. Relative errors are calculated as the ratio of estimation error to the trend estimate in the counterfactual scenario. MLR (5 features) only use temperature, precipitation, humidity, and surface wind speed (U,V directions) as the meteorological features.

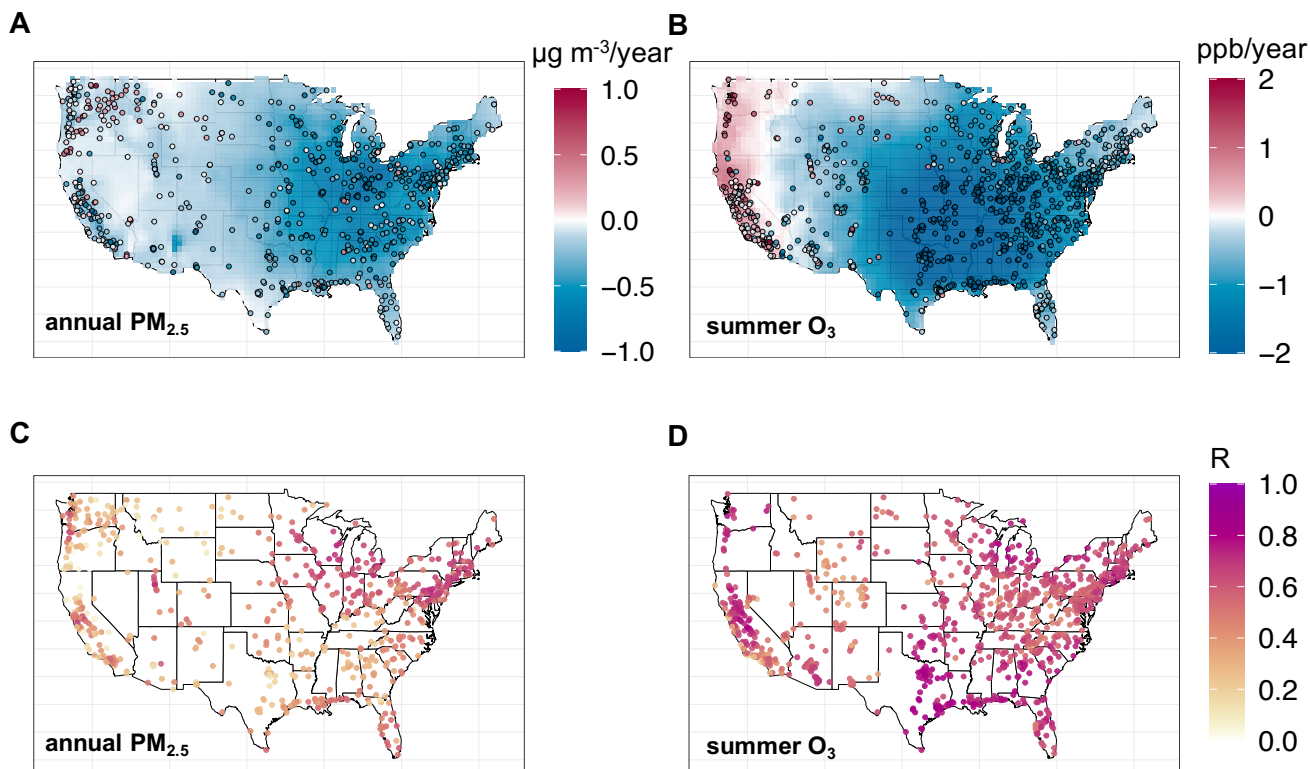



Figure A1. Comparison between the annual PM$_{2.5}$ (Panels A and C) and summer O$_3$ (Panels B and D) concentrations measured by the monitoring network and GEOS-Chem simulations in the US (2011-2017). Panels A and B show the trends in monitored concentrations (dots) and trends in the observational scenarios in GEOS-Chem simulations (background) without meteorology corrections. Panels C and D show the Pearson correlation coefficient (R) between the daily measured concentrations and simulated concentrations. Observational air quality data is derived from U.S. Environmental Protection Agency (2021c).

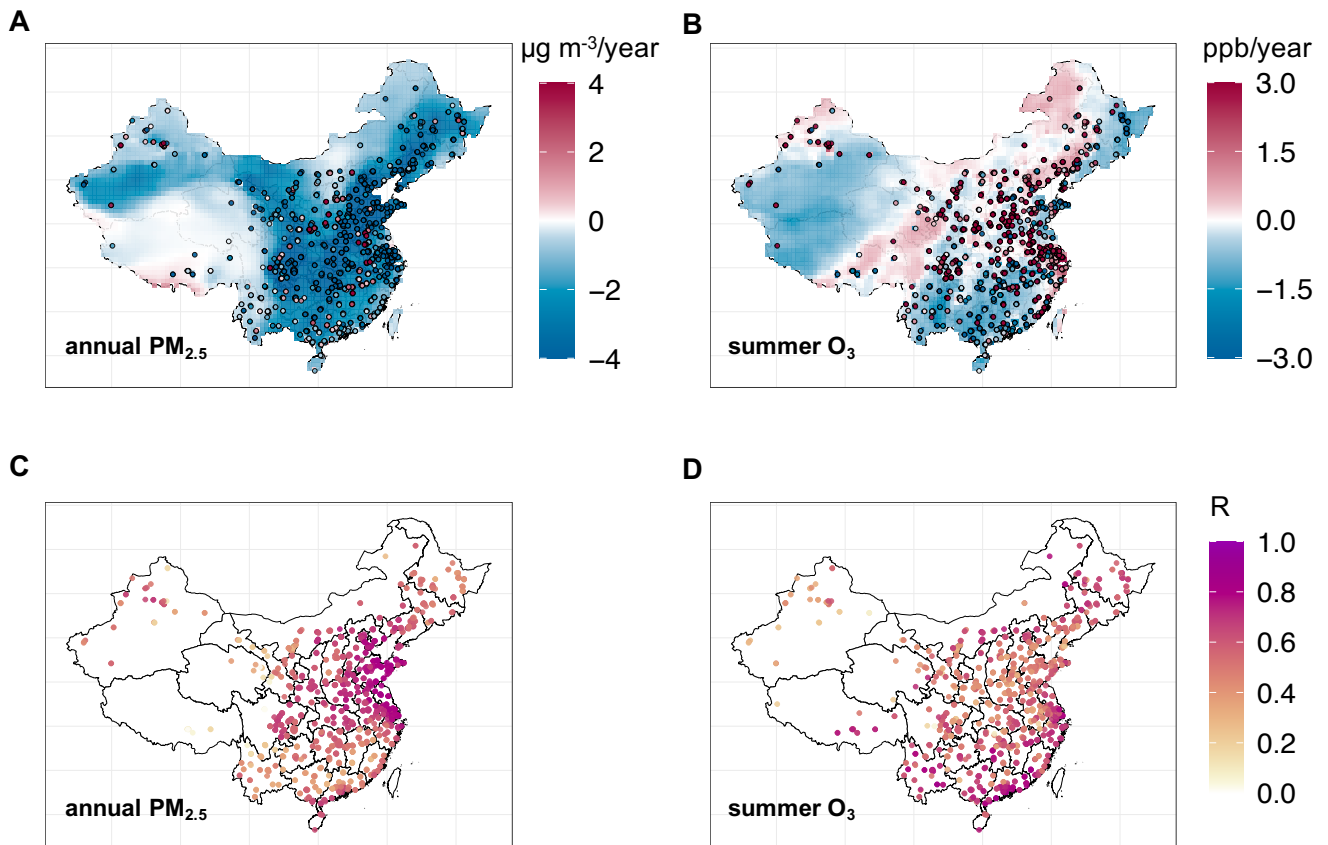



Figure A2. Comparison between the annual PM$_{2.5}$ (Panels A and C) and summer O$_3$ (Panels B and D) concentrations measured by the surface monitoring network and GEOS-Chem simulations in China (2014-2017). Panels A and B show the trends in monitored concentrations (dots) and trends in the observational scenarios in GEOS-Chem simulations (background) without meteorology corrections. Panels C and D show the Pearson correlation coefficient (R) between the daily measured concentrations and simulated concentrations. Observational air quality data is derived from China's Ministry of Ecology and Environment (2021).

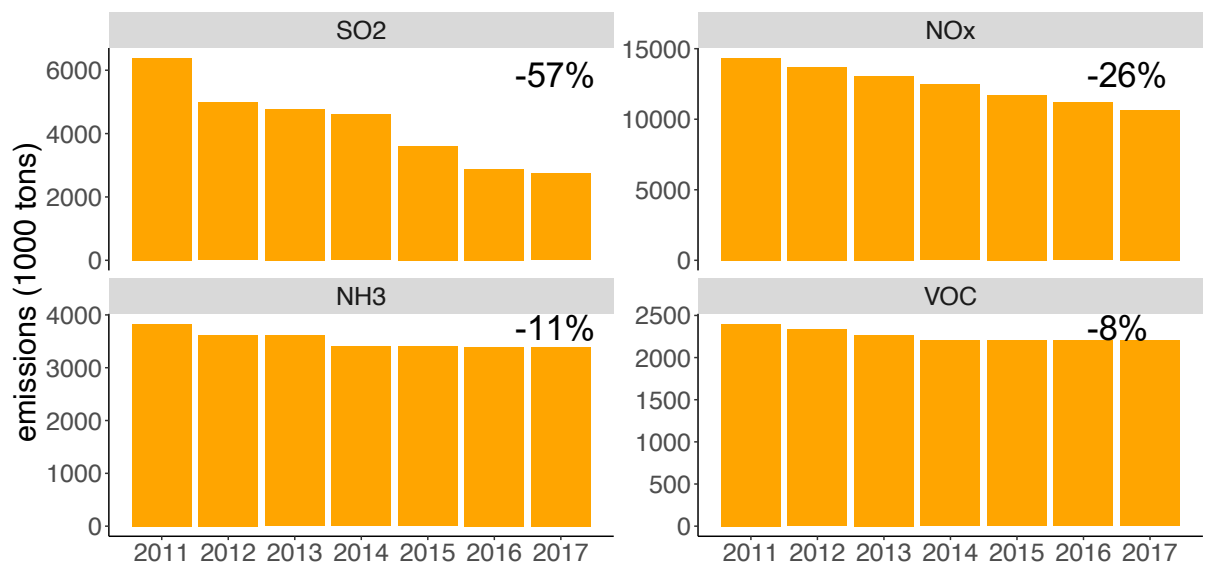



Figure A3. National total anthropogenic emissions in the US (2011- 2017). The emissions data is derived from the national total emissions of criteria air pollutants reported by the US EPA Air Emissions Inventory (U.S. Environmental Protection Agency, 2021a). Changes in emissions between 2011 and 2017 as percentages of the emissions in 2011 are presented in the figure.

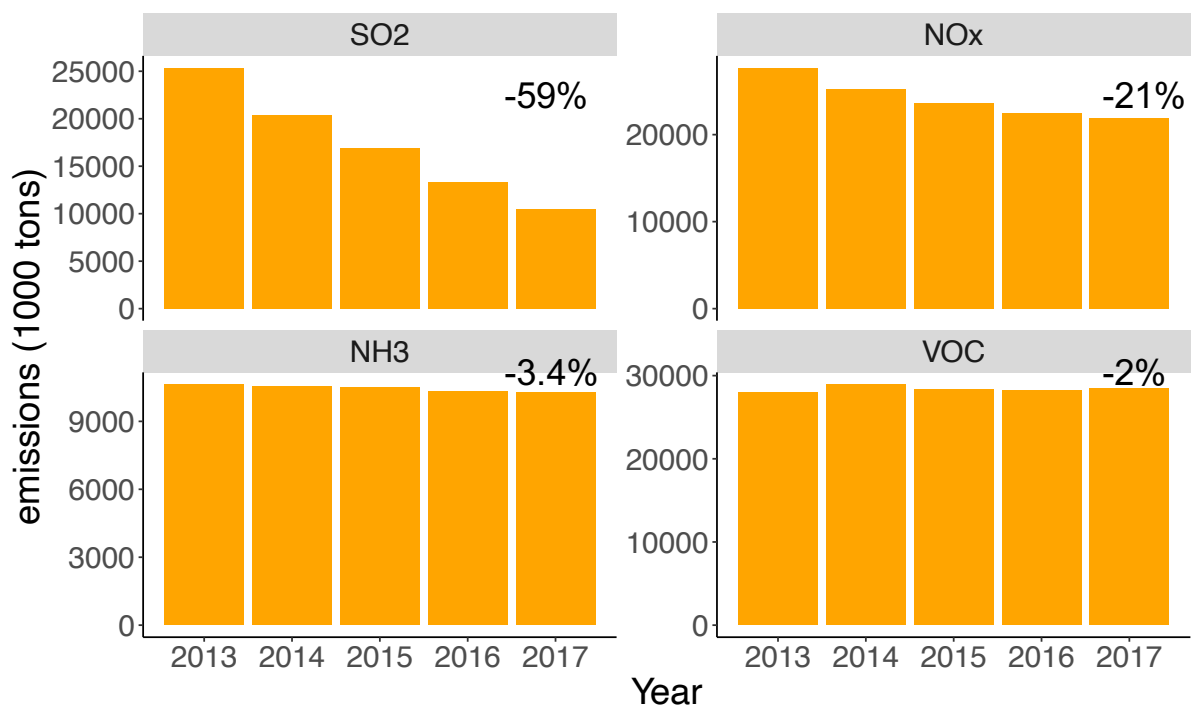



Figure A4. National total anthropogenic emissions in China (2013- 2017). The emissions data is derived from the Multi-resolution Emission Inventory (MEIC) (Li et al., 2017b). Changes in emissions between 2013 and 2017 as percentages of the emissions in 2013 are presented in the figure.

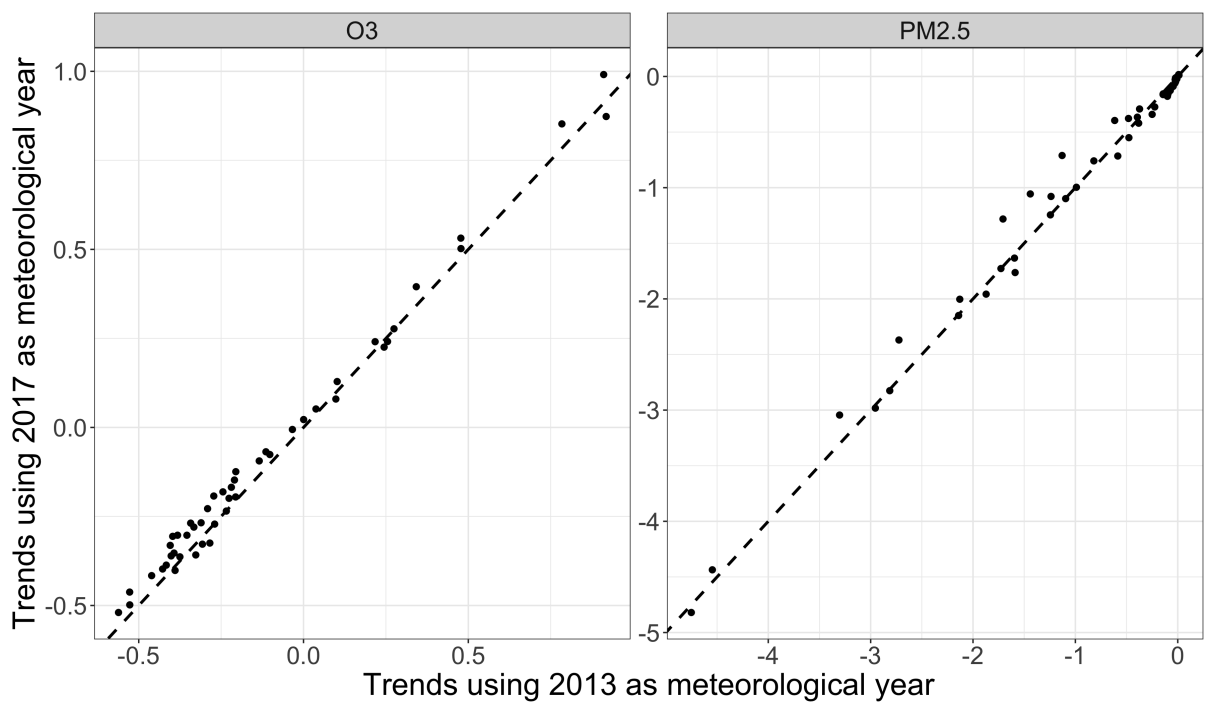



Figure A5. Counterfactual trends in O_3 (unit: ppb/yr) and $PM_{2.5}$ (unit: $\mu g\ m^{-3}$/year) in China with different meteorological years. Each dot represents one grid cell in China. The x-axis shows the trends in air quality in the counterfactual scenario using the meteorological field in 2013, and the Y-axis shows the trends in air quality in the counterfactual scenario using the meteorological field in 2017. Results here are derived from simulation at 4×5 degrees.

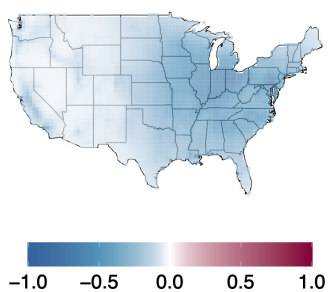
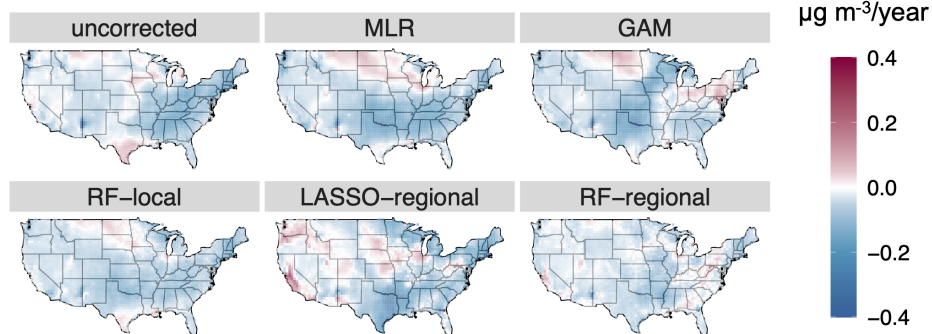
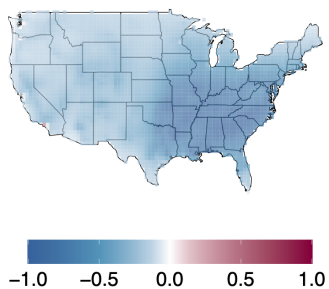
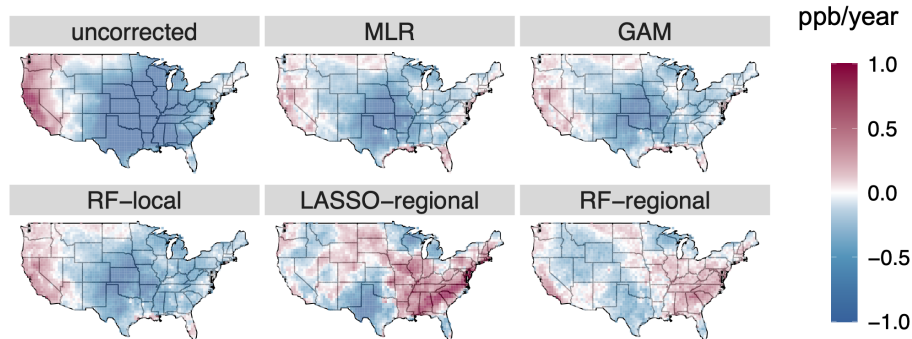



Figure A6. Trend estimates of daily annual PM$_{2.5}$ (Panels A and B) and summer O$_3$ (C and D) in the US. Panels A and C show trend estimates under the counterfactual scenario (β^{count}). Panels B and D show the estimation errors of trend estimates under different correction methods compared with the counterfactual scenarios ($\beta^{obs} - \beta^{count}$). The average of the absolute error for each method is shown in the figure. Unit of trend estimate is μg m^{-3}/year for PM$_{2.5}$ or ppb/year for O$_3$.

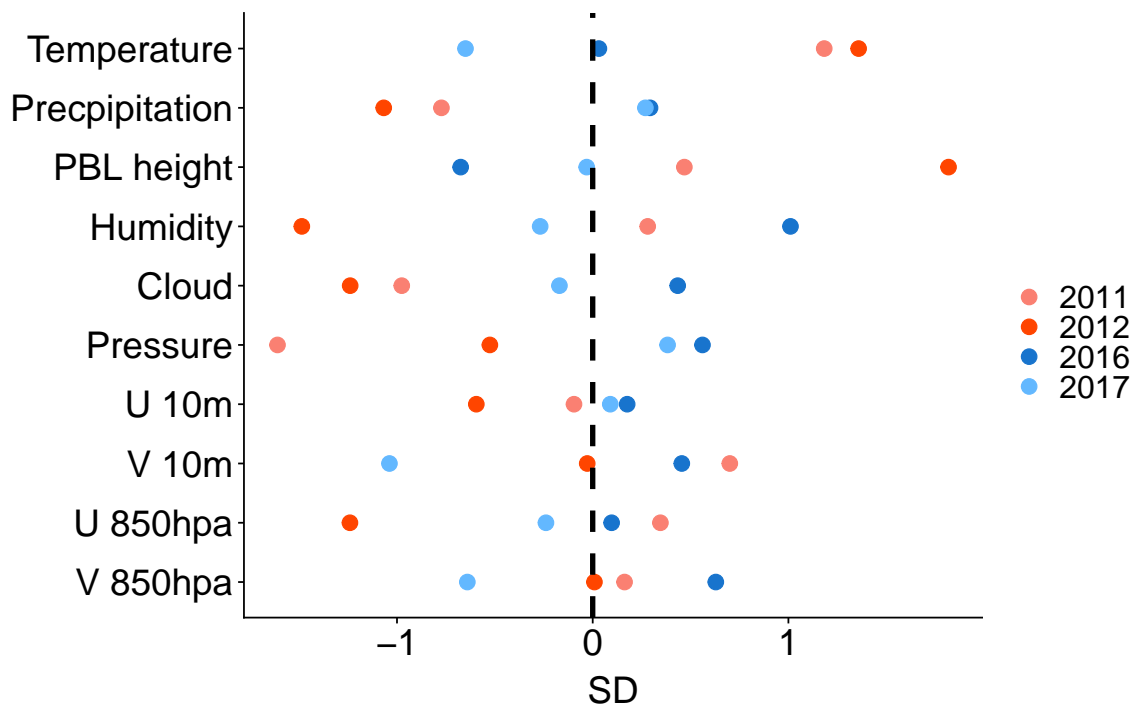



Figure A7. Deviations of meteorological features from the 7-year average in the US (South and Midwest). The deviation is quantified in the units of standard deviation (SD) across the 7-year period. Zero indicates the 7-year average. This plot shows the summer time average of daily MDA8 meteorological variables for each year aggregated over South and Midwest US.

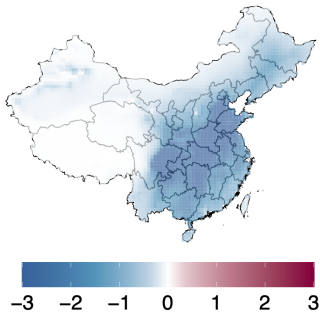
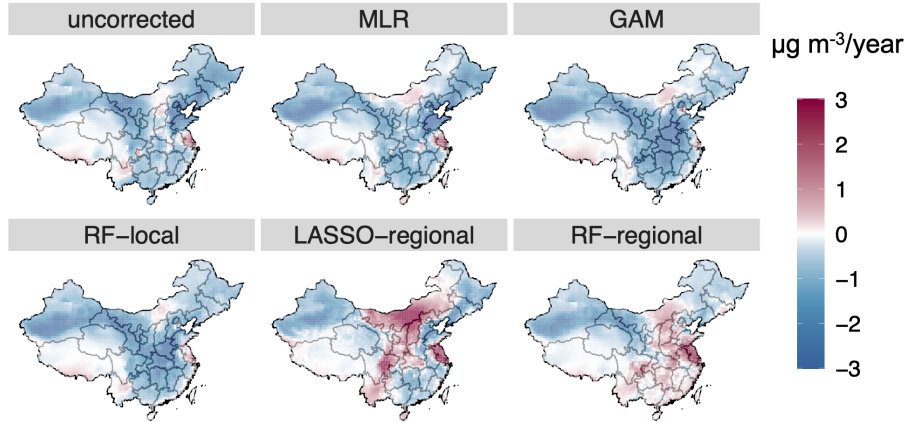
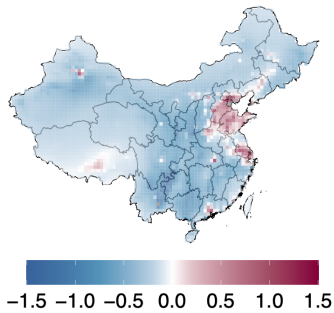
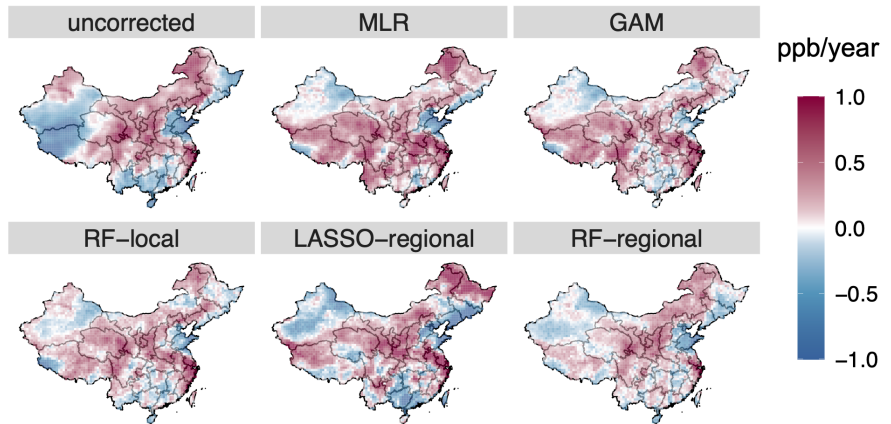



Figure A8. Trend estimates of daily annual PM$_{2.5}$ (Panels A and B) and summer O$_3$ (C and D) in China. Panels A and C show trend estimates under the counterfactual scenario (β^{count}). Panels B and D show the estimation errors of trend estimates under different correction methods compared with the counterfactual scenarios ($\beta^{obs} - \beta^{count}$). The average of the absolute error for each method is shown in the figure. Unit of trend estimate is μg m^{-3}/year for PM$_{2.5}$ or ppb/year for O$_3$.

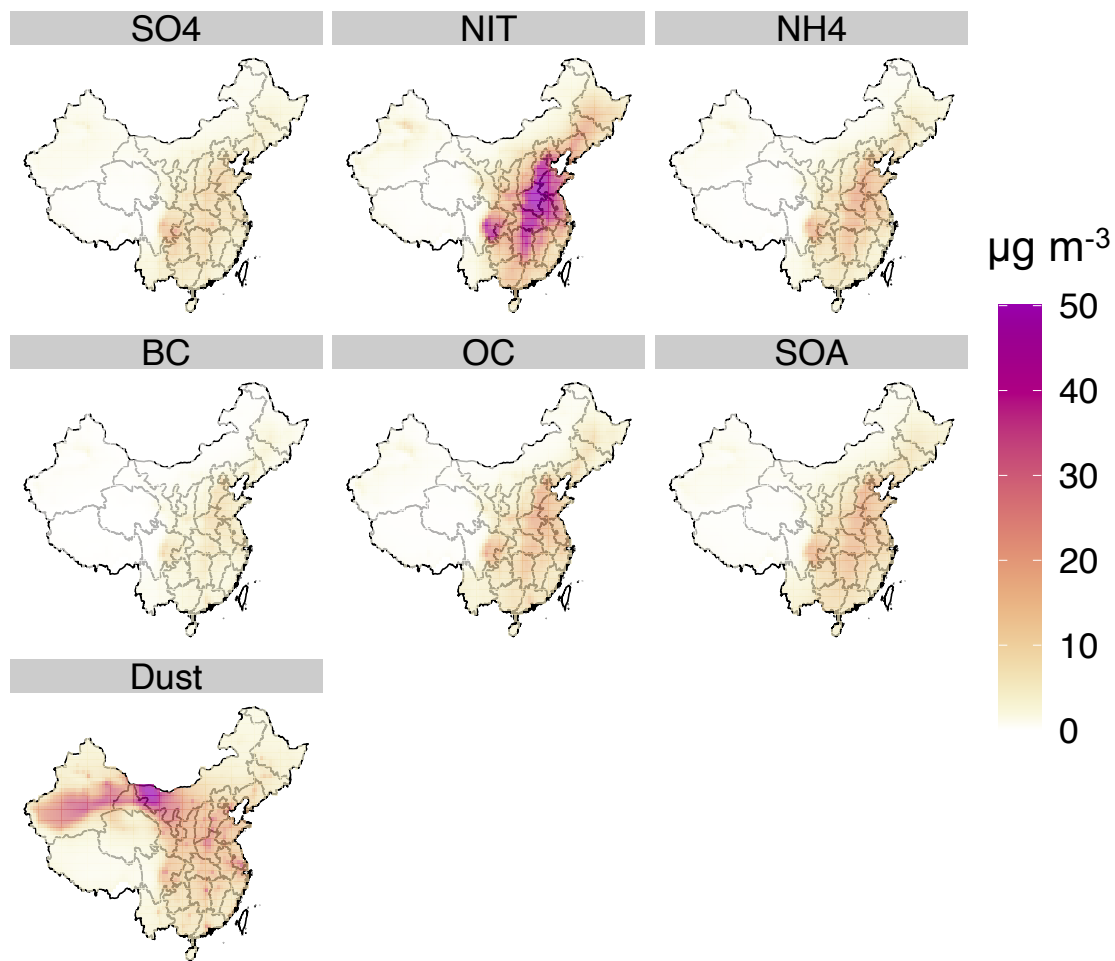



Figure A9. Concentrations of component species of PM$_{2.5}$ in China (average across 2013-2017). The figure shows concentrations of sulfate (SO$_4$), nitrate (NIT), ammonium (NH$_4$), black carbon (BC), organic carbon (OC), secondary organic aerosol (SOA), and dust.

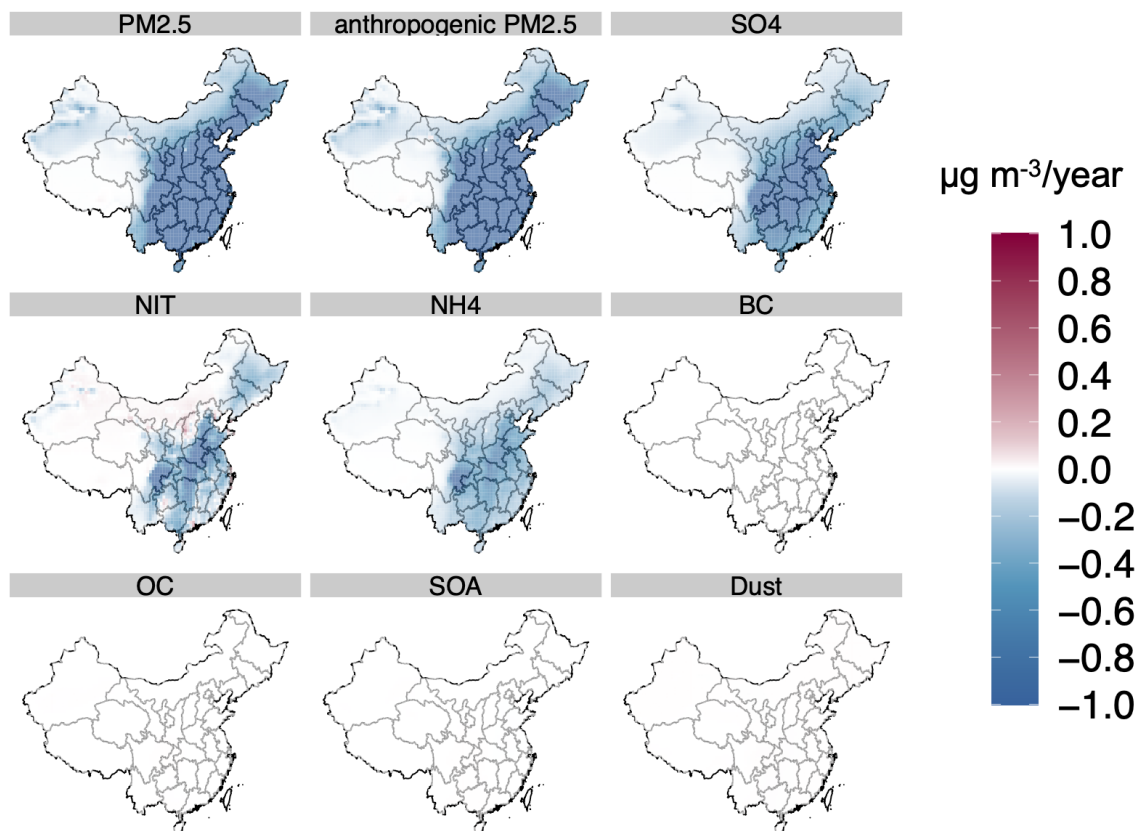



Figure A10. Counterfactual trends of component species of PM$_{2.5}$ in China. The figure shows counterfactual trends of total PM$_{2.5}$, anthropogenic PM$_{2.5}$ (total PM$_{2.5}$ excluding dust and sea salt), sulfate (SO$_4$), nitrate (NIT), ammonium (NH$_4$), black carbon (BC), organic carbon (OC), secondary organic aerosol (SOA), and dust.

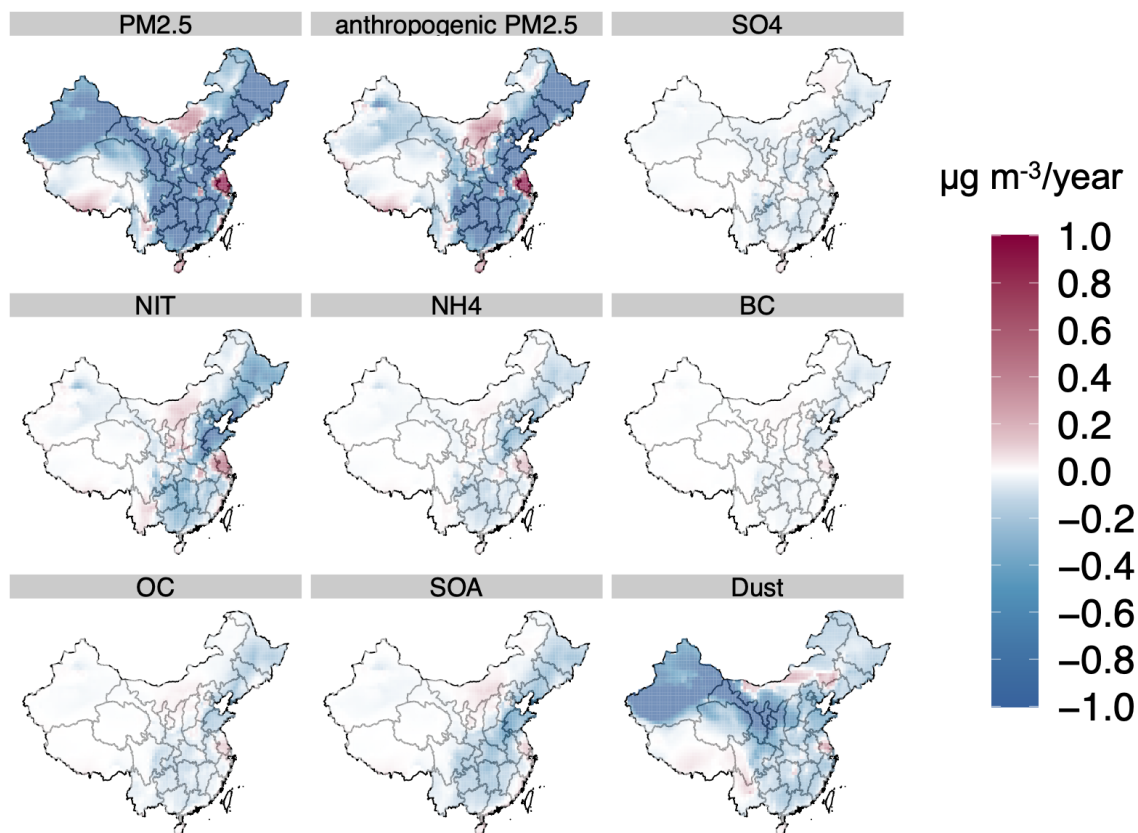



Figure A11. Differences between counterfactual trends and trends evaluated under MLR ($\beta^{MLR} - \beta^{count}$) of component species of PM$_{2.5}$ in China. The figure shows estimation errors of total PM$_{2.5}$, anthropogenic PM$_{2.5}$ (total PM$_{2.5}$ excluding dust and sea salt), sulfate (SO$_4$), nitrate (NIT), ammonium (NH$_4$), black carbon (BC), organic carbon (OC), secondary organic aerosol (SOA) and dust.

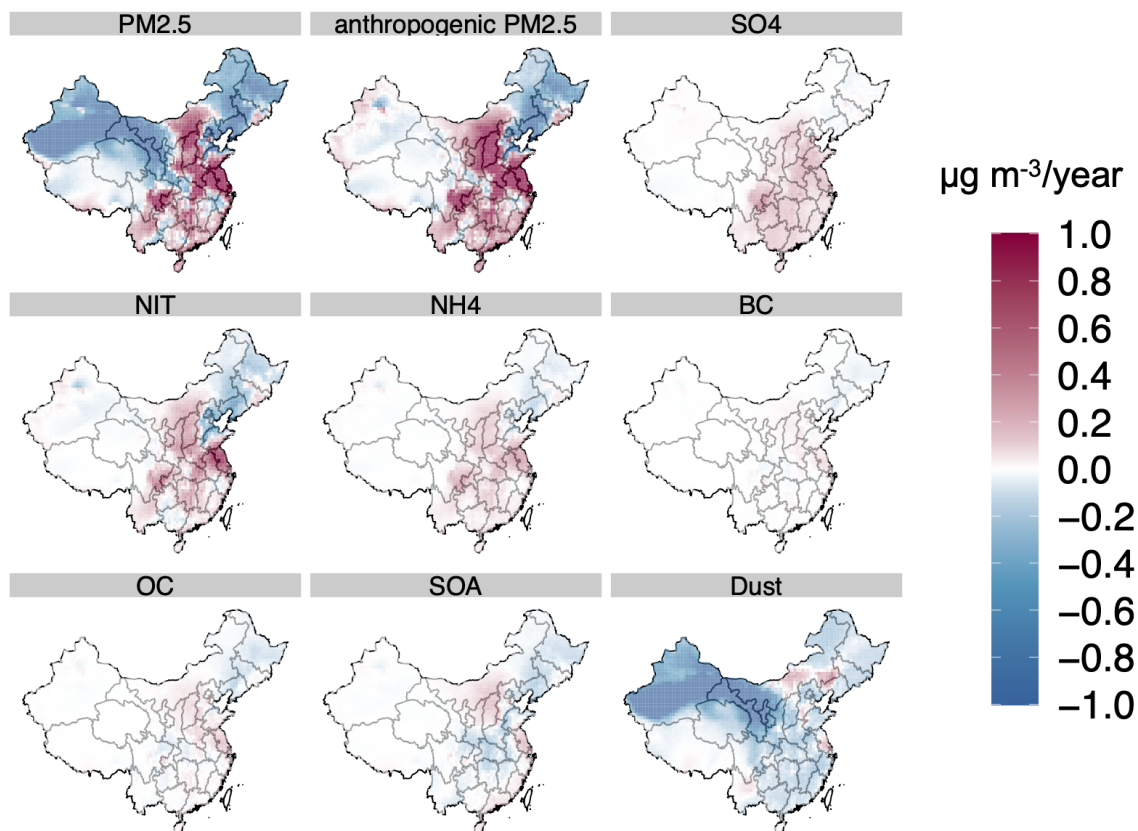



Figure A12. Differences between counterfactual trends and trends evaluated under RF-regional ($\beta^{RF-regional} - \beta^{count}$) of component species of PM$_{2.5}$ in China. The figure shows estimation errors of total PM$_{2.5}$, anthropogenic PM$_{2.5}$ (total PM$_{2.5}$ excluding dust and sea salt), sulfate (SO$_4$), nitrate (NIT), ammonium (NH$_4$), black carbon (BC), organic carbon (OC), secondary organic aerosol (SOA) and dust.

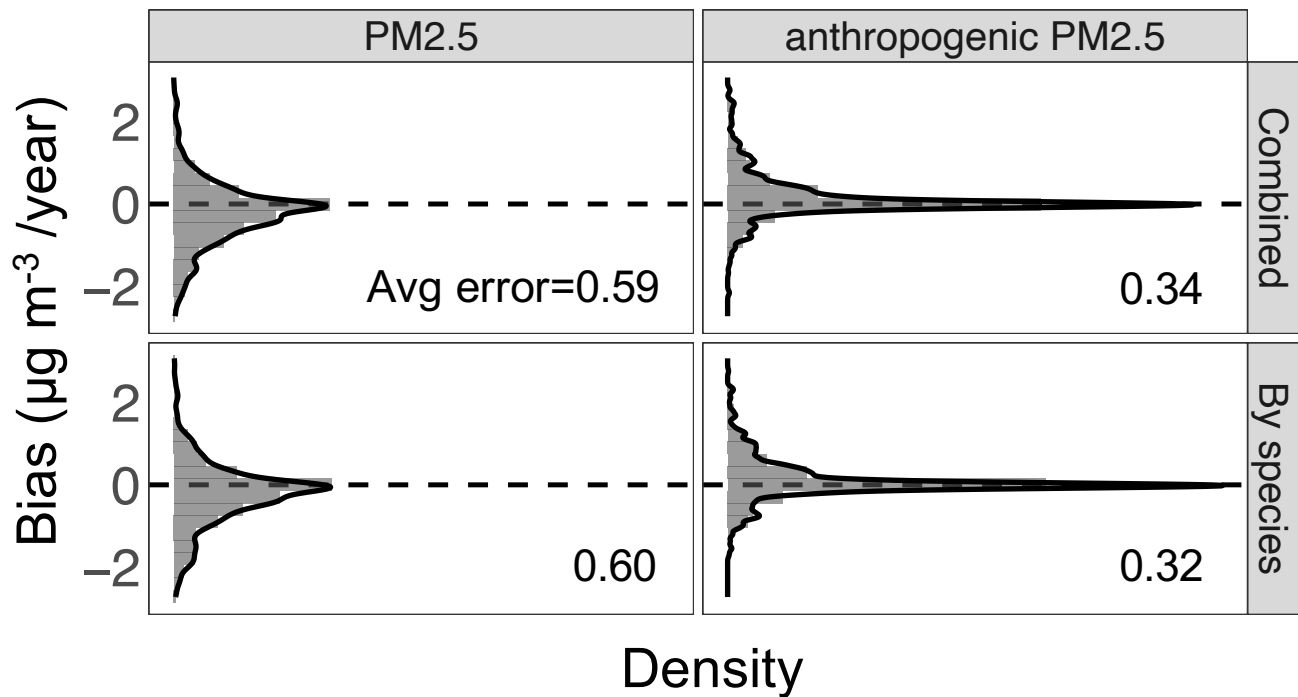
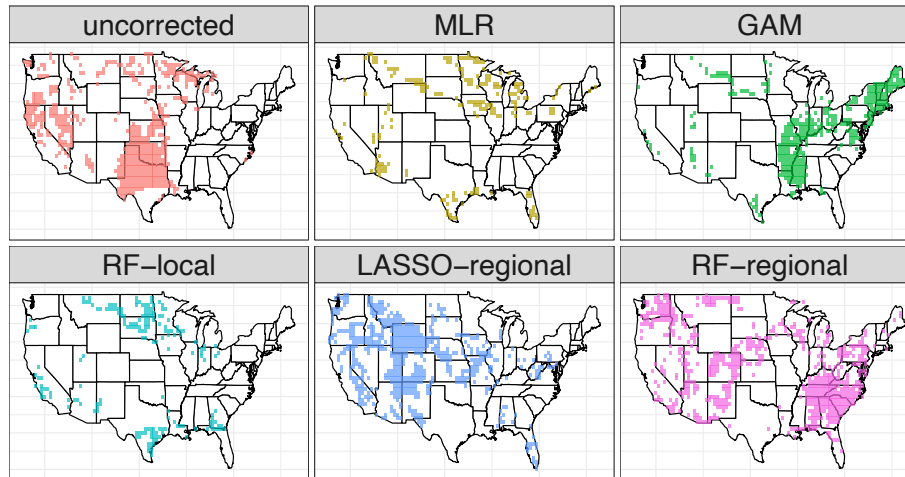



Figure A13. Histograms of estimation errors of trend estimates in $PM_{2.5}$ under two implementations of the *RF-regional* method (applied to China). The upper panels (*Combined*) show results of fitting RF models to the total concentrations of $PM_{2.5}$ to directly estimate trends (the main results). The lower panels (*By species*) show results of fitting RF models to individual $PM_{2.5}$ species and then combine predictions to estimate trends. The left panels show results for total $PM_{2.5}$ and right panels show results for anthropogenic $PM_{2.5}$ (total $PM_{2.5}$ excluding dust and sea salt). Average of the estimation errors across all grid cells is shown in the figure.

A   annual PM$_{2.5}$

B   summer O$_3$

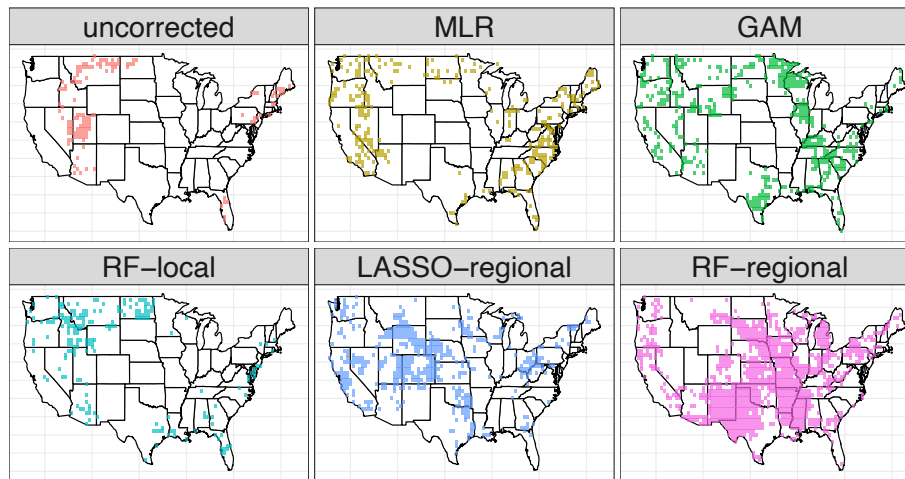
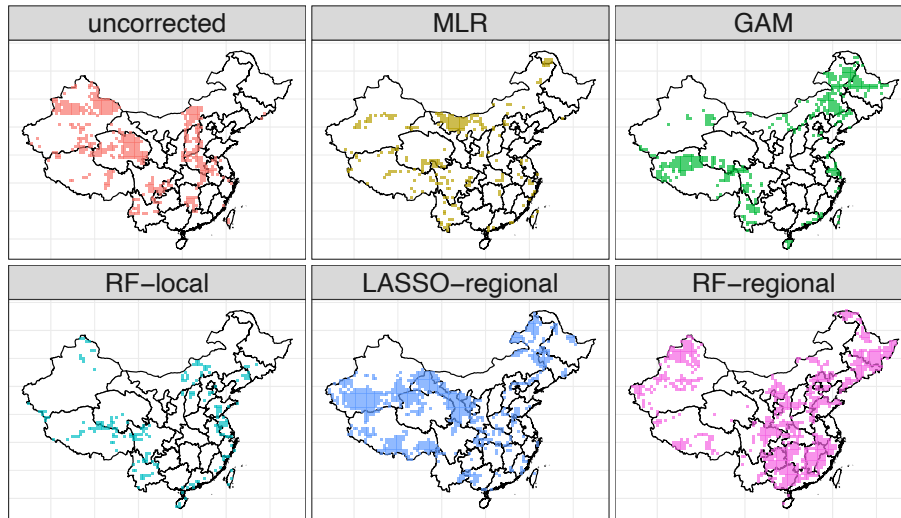



Figure A14. Best-performing correction method for each grid cell (US). For each method, the figure shows the grid cells at which the trend estimate has the smallest estimation error (i.e. closest to the trend in the counterfactual scenario) among the tested methods.

A annual PM$_{2.5}$

B summer O$_3$

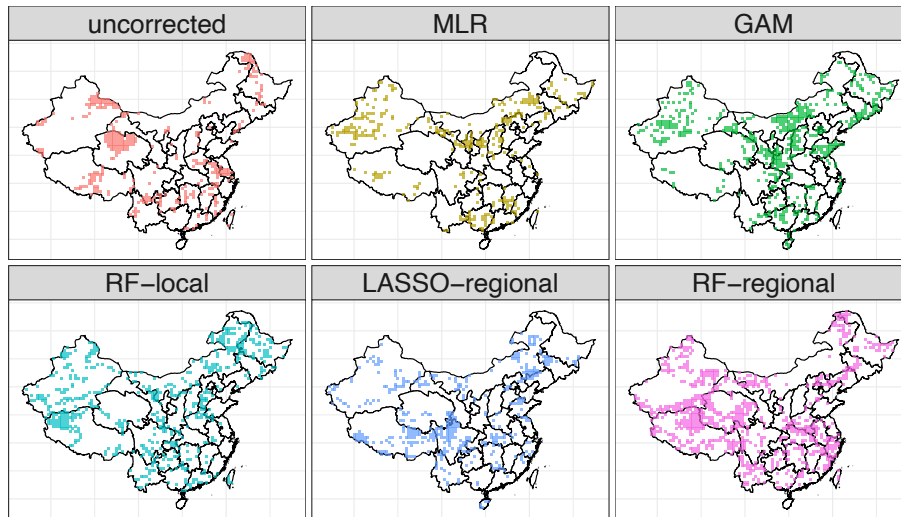



Figure A15. Best-performing correction method for each grid cell (China). For each method, the figure shows the grid cells at which the trend estimate has the smallest estimation error (i.e. closest to the trend in the counterfactual scenario) among the tested methods.

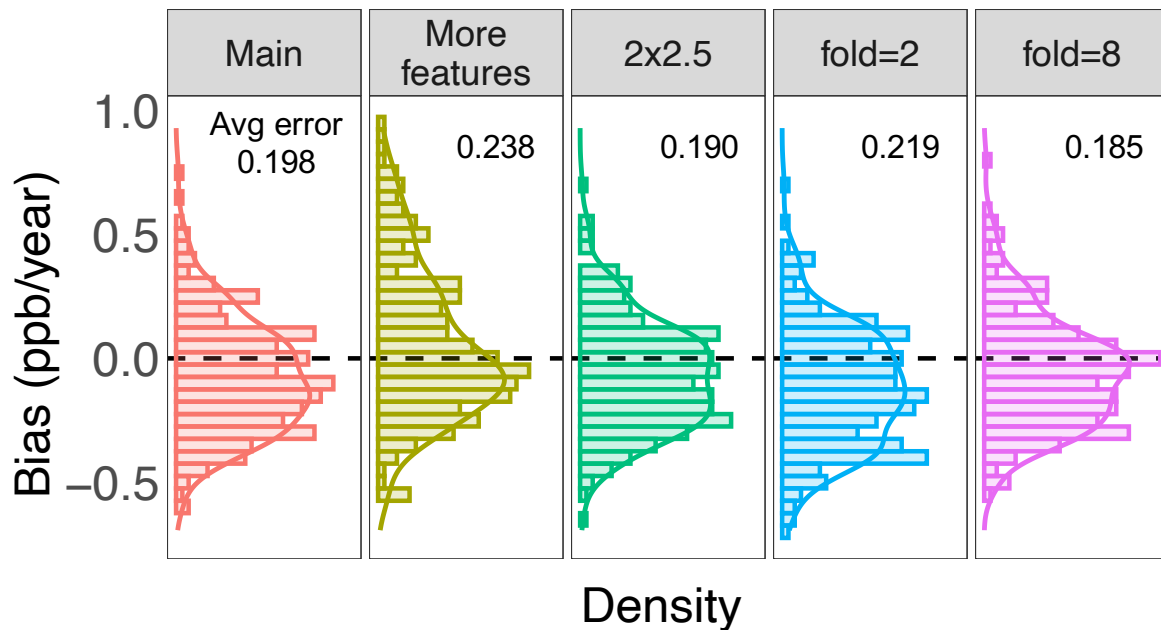



Figure A16. Histograms of estimation errors of trend estimates in O_3 under different implementations of the *RF-regional* method (applied to the US). From left to right: Main (the main results), More features (including 9 extra meteorological features), 2x2.5 (using regional features with spatial resolution at $2\times2.5°$, instead of $4\times5°$), fold=2 (using 2 folds for data-splitting and cross-fitting), fold=8 (using 8 folds for data-splitting and cross-fitting). Average of the absolute error for each implementation is shown in the figure. Here we only use a random subset of all the grid cells in the US due to high computational cost.

References

Abatzoglou, J. T. and Williams, A. P.: Impact of anthropogenic climate change on wildfire across western US forests, Proceedings of the National Academy of Sciences, 113, 11 770–11 775, 2016.

Beijing Municipal Ecology and Environment Bureau: Beijing Clean Air Action Plan (2013–2017), 744, 140 837, http://sthjj.beijing.gov.cn/bjhrb/index/xxgk69/sthjlyzwg/wrygl/603133/index.html, 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, Journal of Geophysical Research Atmospheres, 106, 23 073–23 095, 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, Proceedings of the National Academy of Sciences, 118, 2021.

Camalier, L., Cox, W., and Dolwick, P.: The effects of meteorology on ozone in urban areas and their use in assessing ozone trends, Atmospheric Environment, 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, Atmospheric Environment, 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 PM2. 5 and O3 trends and associated health burden since China's clean air actions, Science of The Total Environment, 744, 140 837, 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 PM2.5 concentrations in Beijing from 2013 to 2017, Atmospheric Chemistry and Physics, 19, 13 519–13 533, 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 PM2.5 air quality improvement in Beijing during 2013-2017: A model-based decomposition analysis, Atmospheric Chemistry and Physics, 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, 2018.

China's Ministry of Ecology and Environment: National Air Quality Monitoring Data, https://quotsoft.net/air/, 2021.

European Union: Air Quality Standards in the European Union, 744, 140 837, https://ec.europa.eu/environment/air/quality/standards.htm, 2020.

Fiore, A. M., Dentener, F., Wild, O., Cuvelier, C., Schultz, M., Hess, P., Textor, C., Schulz, M., Doherty, R., Horowitz, L., et al.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, Journal of Geophysical Research: Atmospheres, 114, 2009.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of climate, 30, 5419–5454, 2017.

538 Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and Hueglin, C.: Random forest meteorological normalisation models for Swiss

539   PM10 trend analysis, Atmospheric Chemistry and Physics, 18, 6223–6239, https://doi.org/10.5194/acp-18-6223-2018, 2018.

540 Han, H., Liu, J., Shu, L., Wang, T., and Yuan, H.: Local and synoptic meteorological influences on daily variability in summertime surface

541   ozone in eastern China, Atmospheric Chemistry and Physics, 20, 203–222, https://doi.org/10.5194/acp-20-203-2020, 2020.

542 Hayn, M., Beirle, S., Hamprecht, F. A., Platt, U., Menze, B. H., and Wagner, T.: Analysing spatio-temporal patterns of the global NO2-

543   distribution retrieved from GOME satellite observations using a generalized additive model, Atmospheric Chemistry and Physics, 9,

544   6459–6477, https://doi.org/10.5194/acp-9-6459-2009, 2009.

545 He, P., Liang, J., Qiu, Y. L., Li, Q., and Xing, B.: Increase in domestic electricity consumption from particulate air pollution, Nature Energy,

546   5, 985–995, 2020.

547 Heald, C. L., Collett, J. L., Lee, T., Benedict, K. B., Schwandner, F. M., Li, Y., Clarisse, L., Hurtmans, D. R., Van Damme, M., Clerbaux, C.,

548   Coheur, P. F., Philip, S., Martin, R. V., and Pye, H. O.: Atmospheric ammonia and particulate inorganic nitrogen over the United States,

549   Atmospheric Chemistry and Physics, 12, 10 295–10 312, https://doi.org/10.5194/acp-12-10295-2012, 2012.

550 Henneman, L. R., Holmes, H. A., Mulholland, J. A., and Russell, A. G.: Meteorological detrending of primary and secondary pollutant

551   concentrations: Method application and evaluation using long-term (2000-2012) data in Atlanta, Atmospheric Environment, 119, 201–

552   210, https://doi.org/10.1016/j.atmosenv.2015.08.007, 2015.

553 Holland, D. M., Principe, P. P., and Sickles, J. E.: Trends in atmospheric sulfur and nitrogen species in the eastern United States for 1989-

554   1995, Atmospheric Environment, 33, 37–49, https://doi.org/10.1016/S1352-2310(98)00123-X, 1998.

555 Keller, C. A., Long, M. S., Yantosca, R. M., Da Silva, A., Pawson, S., and Jacob, D. J.: HEMCO v1. 0: a versatile, ESMF-compliant

556   component for calculating emissions in atmospheric models, Geoscientific Model Development, 7, 1409–1417, 2014.

557 Leung, D. M., Tai, A. P., Mickley, L. J., Moch, J. M., Van Donkelaar, A., Shen, L., and Martin, R. V.: Synoptic meteorological modes of

558   variability for fine particulate matter (PM2.5) air quality in major metropolitan regions of China, Atmospheric Chemistry and Physics, 18,

559   6733–6748, https://doi.org/10.5194/acp-18-6733-2018, 2018.

560 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,

561   M., Brauer, M., and Zhang, Q.: Trends in Chemical Composition of Global and Regional Population-Weighted Fine Particulate Matter

562   Estimated for 25 Years, Environmental Science and Technology, 51, 11 185–11 195, https://doi.org/10.1021/acs.est.7b02530, 2017a.

563 Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., and Bates, K. H.: Anthropogenic drivers of 2013-2017 trends in summer

564   surface ozone in China., Proceedings of the National Academy of Sciences of the United States of America, 116, 422–427,

565   https://doi.org/10.1073/pnas.1812168116, 2018.

566 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: anthro-

567   pogenic and meteorological influences, Atmospheric Chemistry and Physics, 20, 11 423–11 433, 2020.

568 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

569   inventories in China: A review, National Science Review, 4, 834–866, https://doi.org/10.1093/nsr/nwx150, 2017b.

Lu, X., Zhang, L., Chen, Y., Zhou, M., Zheng, B., Li, K., Liu, Y., Lin, J., Fu, T. M., and Zhang, Q.: Exploring 2016-2017 surface ozone pollution over China: Source contributions and meteorological influences, Atmospheric Chemistry and Physics, 19, 8339–8361, https://doi.org/10.5194/acp-19-8339-2019, 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, Environmental Science & Technology Letters, 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, Atmospheric Chemistry and Physics, 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, Proceedings of the National Academy of Sciences of the United States of America, 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., Colette, 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 surface ozone over Europe, Atmospheric Chemistry and Physics, 18, 12 269–12 288, https://doi.org/10.5194/acp-18-12269-2018, 2018.

Porter, W. C. and Heald, C. L.: The mechanisms and meteorological drivers of the ozone–temperature relationship, Atmospheric Chemistry and Physics Discussions, pp. 1–26, https://doi.org/10.5194/acp-2019-140, 2019.

Qu, L., Liu, S., Ma, L., Zhang, Z., Du, J., Zhou, Y., and Meng, F.: Evaluating the meteorological normalized PM2.5 trend (2014–2019) in the "2+26" region of China using an ensemble learning technique, Environmental Pollution, 266, 115 346, 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, Nature Communications, 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, Environmental Science and Technology, 53, 1098–1108, https://doi.org/10.1021/acs.est.8b05094, 2019.

Shen, L., Mickley, L. J., and Tai, A. P.: Influence of synoptic patterns on surface ozone variability over the eastern United States from 1980 to 2012, Atmospheric Chemistry and Physics, 15, 10 925–10 938, 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, Atmospheric Chemistry and Physics, 16, 12 239–12 271, 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, Science Advances, 7, 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), 744, 140 837, http://www.gov.cn/zwgk/2013-09/12/content_2486773.htm, 2013.

Tai, A. P., Mickley, L. J., Jacob, D. J., Leibensperger, E. M., Zhang, L., Fisher, J. A., and Pye, H. O.: Meteorological modes of variability for fine particulate matter (PM2.5) air quality in the United States: Implications for PM2.5 sensitivity to climate change, Atmospheric Chemistry and Physics, 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, Atmospheric Chemistry and Physics, 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., Crounse, 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?, Atmospheric Chemistry and Physics, 16, 13 561–13 577, https://doi.org/10.5194/acp-16-13561-2016, 2016.

U.S. Energy Information Agency: Heat wave results in highest U.S. electricity demand since 2017, 744, 140 837, https://www.eia.gov/todayinenergy/detail.php?id=40253, 2019.

U.S. Environmental Protection Agency: NATIONAL PRIMARY AND SECONDARY AMBIENT AIR QUALITY STANDARDS, 744, 140 837, https://ecfr.federalregister.gov/current/title-40/chapter-I/subchapter-C/part-50, 2019.

U.S. Environmental Protection Agency: National Emissions Inventory 2017: Technical Support Document, https://www.epa.gov/sites/production/files/2021-02/documents/nei2017_tsd_full_jan2021.pdf, 2021a.

U.S. Environmental Protection Agency: Criteria pollutants National Tier 1 for 1970 - 2020, 744, 140 837, https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data, 2021b.

U.S. Environmental Protection Agency: Air Data: Air Quality Data Collected at Outdoor Monitors Across the US, https://aqs.epa.gov/aqsweb/airdata/download_files.html#Meta/, 2021c.

Vu, T., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., and Harrison, R.: Assessing the impact of Clean Air Action Plan on Air Quality Trends in Beijing Megacity using a machine learning technique, Atmospheric Chemistry and Physics, pp. 1–18, https://doi.org/10.5194/acp-2019-173, 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, Journal of Geophysical Research: Atmospheres, 109, 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 interannual variability in meteorological conditions, Atmospheric Environment, 248, 118 234, 2021.

Werf, G. R., Randerson, J. T., Giglio, L., Leeuwen, T. T. v., Chen, Y., Rogers, B. M., Mu, M., Van Marle, M. J., Morton, D. C., Collatz, G. J., et al.: Global fire emissions estimates during 1997–2016, Earth System Science Data, 9, 697–720, 2017.

Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, Journal of the Royal Statistical Society (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, Journal of Geophysical Research: Atmospheres, 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, Proceedings of the National Academy of Sciences, 119, e2111372 119, 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, Atmospheric Chemistry and Physics, 19, 11 031–11 041, 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, Nature Geoscience, 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 24h-average PM2.5 concentrations in the North China Plain during 2013–2017, Science of the Total Environment, 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 PM2.5 air quality in China from 2013 to 2017, Proceedings of the National Academy of Sciences of the United States of America, pp. 1–7, 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, Environmental Science and Technology, 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, Environmental Science Technology Letters, 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., et al.: Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions, Atmospheric Chemistry and Physics, 18, 14 095–14 111, 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 PM2.5 Concentrations and Meteorological Factor-Induced Fluctuations, Environmental Science and Technology, 52, 10 416–10 425, 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.