https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

Source specific bias correction of US background ozone modeled in CMAQ

T. Nash Skipper[1], Christian Hogrefe[2], Barron H. Henderson[2], Rohit Mathur[2], Kristen M. Foley[2], Armistead G. Russell[1]

[1]School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2]U.S. Environmental Protection Agency, Research Triangle Park, NC, 27709, USA

Correspondence to: Armistead G. Russell (ar70@gatech.edu)

Abstract. United States (US) background ozone (O_3) is the counterfactual O_3 that would exist with zero US anthropogenic emissions. Estimates of US background O_3 typically come from chemical transport models (CTMs), but different models vary in their estimates of both background and total O_3. Here, a measurement-model data fusion approach is used to estimate CTM biases in US anthropogenic O_3 and multiple US background O_3 sources, including natural emissions, long-range international emissions, short-range international emissions from Canada and Mexico, and stratospheric O_3. Spatially and temporally varying bias correction factors adjust each simulated O_3 component so that the sum of the adjusted components evaluates better against observations compared to unadjusted estimates. The estimated correction factors suggest a seasonally consistent positive bias in US anthropogenic O_3 in the eastern US, with the bias becoming higher with coarser model resolution and with higher simulated total O_3 though the bias does not increase much with higher observed O_3. Correlation among different US background O_3 components can increase the uncertainty in the estimation of the source-specific adjustment factors. Despite this, results indicate that there may be a negative bias in modeled estimates of the impact of stratospheric O_3 at the surface. This type of data fusion approach can be extended to include data from multiple models to leverage the strengths of different data sources while reducing uncertainty in the US background ozone estimates.

1 Introduction

United States (US) background (USB) ozone (O_3) is the counterfactual O_3 that would exist if US anthropogenic (USA) emissions were zero. The National Ambient Air Quality Standard (NAAQS) for O_3

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

was set at a level of 70 ppb in 2015 and may be lowered. In its recent reviews of the O_3 NAAQS, the US Environmental Protection Agency (EPA) noted the importance of USB O_3 (USEPA, 2013, 2014, 2020a, b). USB O_3 takes up a larger portion of the allowed ozone as the NAAQS is tightened and is a larger portion of total observed O_3 as anthropogenic precursor emissions decline. USB O_3 cannot be observed (USEPA, 2013; Fiore et al., 2003; 2010; Mcdonald-Buller et al., 2011; Fiore et al., 2014; Jaffe et al., 2018; USEPA, 2020a, 2014, 2020b). It is typically quantified using a chemical transport model (CTM), most commonly using the zero-out method in which USA emissions are set to zero. There is much uncertainty in CTM estimates of USB O_3 due to model biases and differences in CTM-estimated USB O_3 among different models (Mcdonald-Buller et al., 2011; Fiore et al., 2014; Jaffe et al., 2018; Dolwick et al., 2015; Guo et al., 2018; Huang et al., 2015). Jaffe et al. (2018) estimated that the typical uncertainty in CTM-estimated seasonal mean USB O_3 is ±10 ppb.

Sources of USB O_3 include naturally occurring emissions such as wildfires, biogenic VOCs, oxides of nitrogen (NO_x) from soil, lightning NO_x, stratosphere-to-troposphere exchange, and oxidation of methane. Some portions of total O_3 contributions from soil NO_x and methane oxidation are USB sources while some are anthropogenic. Soil NO_x is emitted by microbial processes in both natural and agricultural lands and is limited by availability of nitrogen in the soil. There is a pre-industrial level of methane that contributes to USB O_3 formation, but any O_3 created through oxidation of methane above the pre-industrial level is anthropogenic. Soil NO_x and methane oxidation are often treated as USB O_3 sources in their entirety in CTM studies due to the complexity of splitting up the natural and anthropogenic portions. USB O_3 sources also include non-US anthropogenic pollution which may be from long range transport or from short range transport from neighboring countries.

In previous work (Skipper et al., 2021), we developed a bias correction method which used regression modeling to adjust CTM-simulated USA and USB O_3 to better align with observations and to improve agreement of differing USB O_3 estimates from different model configurations. We developed spatially and temporally varying scaling factors to adjust USA and USB O_3. In that work, USB O_3 was treated as a single quantity rather than considering different sources of USB O_3 individually. A consistent low bias in USB O_3 in spring was identified, though the specific source of this low bias could not be identified. Here, we extend the bias correction method to estimate biases in separate components of USB

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

O_3. Different sources of USB O_3 are expected to have different seasonal cycles. For example, stratospheric O_3 is expected to peak in the spring while O_3 impacts from natural emissions are expected to peak in summer. Separating the USB O_3 components provides new insights into the inferred CTM error in USB O_3 that was not possible when USB O_3 was treated as a lumped quantity.

2 Methods

2.1 Chemical transport model simulations

Total O_3 (i.e., BASE O_3), USB O_3, and individual USB O_3 components are simulated at both regional and hemispheric scales using the Community Multiscale Air Quality (CMAQ) model. We use maximum daily 8-h average (MDA8) O_3 as the metric of interest since this is the metric used in determining attainment of the NAAQS. References to O_3 throughout are to MDA8 O_3. CMAQ results are from two recent sets of simulations by the US EPA (Table 1). The two sets of simulations include different USB O_3 components allowing us to explore how different components of USB O_3 affect the bias in O_3.

Table 1. Simulation names and descriptions for hemispheric-scale and regional-scale simulations. Table adapted from 2020 O_3 Policy Assessment Table 2-1 (USEPA, 2020b).

Simulation	Description
BASE	All emission sectors are included
ZUSA	All U.S. anthropogenic emissions are removed including prescribed fires.*
ZROW	All international anthropogenic emissions are removed including prescribed fires where possible.**
ZCANMEX	All anthropogenic emissions from Canada and Mexico are removed including prescribed fires where possible.**
ZANTH	All anthropogenic emissions are removed including prescribed fires.**
STRAT	Tracer species for O_3 injected into the upper troposphere/lower stratosphere based on CMAQ potential vorticity parameterization for stratospheric O_3.***

* Emissions estimated to be associated with intentionally set fires (“prescribed fires”) are grouped with anthropogenic fires.
** Only for PA simulations
*** Only for EQUATES simulations.

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

The first set of simulations was conducted for the Policy Assessment (PA) for the review of the O_3 NAAQS in 2020 (USEPA, 2020b). These simulations also support the draft PA for the reconsideration of the O_3 NAAQS. The PA simulations cover the entire year of 2016 and provide estimates of USA and USB O_3 as well as natural (NAT) and international anthropogenic (INTL) contributions to USB O_3. INTL is also further decomposed to short-range international anthropogenic contributions from Canada and Mexico (CANMEX) and long-range international (LINTL) contributions from other countries. The PA simulations consist of nested simulations from hemispheric scale (Mathur et al., 2017) at 108 km horizontal resolution to continental scale at 36 km resolution to a finer continental scale at 12 km resolution.

USB O_3 components are determined by the zero-out method in which the model is run in the same configuration as the base case but with specified emissions sources removed. USB O_3 is estimated by removing US anthropogenic emissions. USA O_3 is calculated as BASE O_3 minus USB O_3. NAT O_3 is estimated by removing all anthropogenic emissions. The non-US anthropogenic O_3 contribution is estimated by removing anthropogenic emissions everywhere except the US. The INTL contribution is calculated as BASE O_3 minus the O_3 with zero anthropogenic emissions in countries other than the US. CANMEX O_3 is estimated by removing Canada and Mexico anthropogenic emissions. The CANMEX O_3 contribution is calculated as BASE O_3 minus the O_3 with zero Canada and Mexico anthropogenic emissions. LINTL O_3 is estimated as INTL O_3 minus CANMEX O_3. Due to non-linear chemistry, there is some residual anthropogenic contribution to BASE O_3 which is not attributed to US or international emissions. Descriptions of these CMAQ simulations and calculation of O_3 components are given in Tables S1 and S2. Further details of the modeling setup are available in the 2020 Policy Assessment (USEPA, 2020b) and are summarized in Tables S4 and S5.

The second set of simulations was developed from EPA's Air QUAlity TimE Series (EQUATES) project which spans 2002-2019. Additional simulations using the EQUATES modeling framework were conducted for 2016–2017 to estimate USB O_3 and USA O_3 using the zero-out method. The EQUATES simulations consist of hemispheric scale simulations at 108 km horizontal resolution and nested US continental scale simulations at 12 km horizontal resolution. Descriptions of these CMAQ simulations and calculation of O_3 components are given in Table S3. Further details on the model configuration for

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

EQUATES are available from Foley et al. (2020) and Foley et al. (2023) and are summarized in Tables S4 and S5.

The 108 km EQUATES simulations also include an inert tracer species which serves as a proxy for simulated stratospheric O_3 contributions. Separate stratospheric O_3 contributions were not available from the PA simulations, so the EQUATES simulations provide an opportunity to assess potential biases specific to stratospheric O_3 contributions. CMAQ simulates stratospheric O_3 using a parameterization based on the relationship between O_3 and potential vorticity (PV) in the upper troposphere and lower stratosphere (UTLS) (Xing et al., 2016). The parameterization was developed using 21 years of ozonesonde data from the World Ozone and Ultraviolet Radiation Data Centre and PV data from the Weather Research Forecasting (WRF) model for 1990-2010. In the EQUATES 108 km simulations, the parameterization is applied to the top model layer only. A PV tracer species tracks O_3 injected into the UTLS throughout the rest of the model domain for the hemispheric simulations. The 12 km continental simulations inherit the PV tracer species through lateral boundary conditions from the hemispheric simulations. This tracer is subject to transport and deposition but not chemistry. We refer to the PV tracer concentrations as STRAT O_3 since it relates to the stratospheric influence, but it only partly replicates the impact of stratospheric O_3 since it does not undergo chemical losses. STRAT O_3 does, however, provide a measure of the spatiotemporal variability of stratospheric O_3 impacts. We also estimate the contribution to USB O_3 from sources other than the stratosphere as USB O_3 minus STRAT O_3 and refer to it as USB_NOSTRAT O_3. The use of the chemically inert PV tracer to split up stratospheric and non-stratospheric influences on USB O_3 introduces uncertainty as the STRAT O_3 component may be unrealistically high, especially in areas and times when there is more active chemistry.

Although the PA and EQUATES simulations both use CMAQ, there are differences in the configuration and inputs used for each set of simulations that may affect simulated O_3 levels (Tables S4 and S5). CMAQ v5.2.1 was used for the PA simulations while CMAQ v5.3.2 was used for the EQUATES simulations. These were the latest versions of CMAQ at the respective times that each set of simulations were conducted. One potential source of differences is that halogen chemistry was updated in CMAQ v5.3 (Sarwar et al., 2019). The EQUATES hemispheric simulations therefore include losses of O_3 over seawater that are not present in the PA hemispheric simulations which could affect O_3 transported over

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

the Pacific in particular. An intercomparison of CMAQ v5.2.1 and CMAQ v5.3.1 (which is not significantly different from CMAQ v5.3.2) showed that the newer version typically had lower O_3 compared to the older version, with mean bias ~1 ppb lower in CMAQ v5.3.1 (Appel et al., 2021). Another significant difference is in the US anthropogenic emissions that were used for the simulations. The PA simulations used an early version (sometimes called the "alpha" version) of a 2016 emissions modeling platform developed by the National Emissions Inventory Collaborative (USEPA, 2019). The EQUATES simulations used an inventory that was developed as part of the broader EQUATES framework to model a long timeseries using consistent methods for emissions estimates (Foley et al., 2023). Differences in the US anthropogenic emissions used in the two model configurations are expected to contribute to differences in simulated O_3.

2.2 O_3 observations

O_3 observational data are from the Air Quality System (AQS) database, which provides data from federal, state, local, and tribal air quality monitoring networks across the US. The average precision of O_3 monitors in the AQS database was reported as 2.2% in 2016 and 2.4% in 2016 and 2017, respectively, and the national average absolute bias was reported as 1.5% in both 2016 and 2017 (https://www.epa.gov/amtic/amtic-ambient-air-monitoring-assessments). There were ~360,000 MDA8 O_3 observations available per year for 2016 and 2017 from ~1250 unique monitoring sites. These numbers take into account monitoring sites where O_3 is measured by multiple instruments at the same site (as indicated in the AQS database by a parameter occurrence code). In these cases, the MDA8 O_3 observations from multiple instruments are averaged for a given site and day and treated as a single observation. The observations overrepresent the eastern US compared to the western US. About 40% of daily MDA8 O_3 observations and ~36% of O_3 monitoring sites are in the western US (as defined by longitude < -97 °W). Western US sites are also overrepresented by sites in the state of California. About 40% of daily MDA8 O_3 observations and ~40% of O_3 monitoring sites in the western US are in California. The observations also overrepresent the high O_3 season of April – October (Figure 1) since many monitors are only required to be operated during the high O_3 season.

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

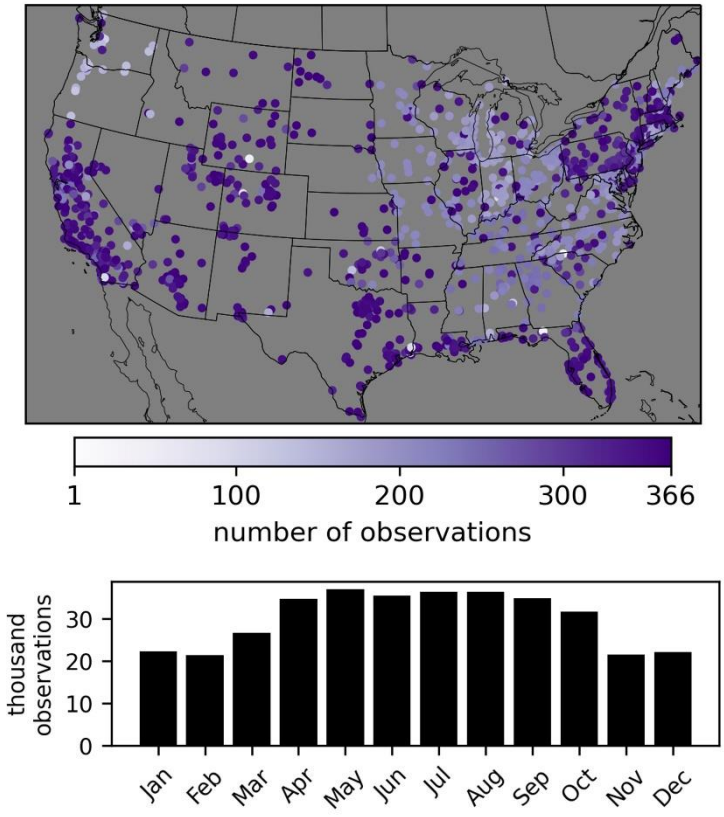


Figure 1. Locations of O_3 observational sites in 2016 indicated with a circle whose color shows the number of daily MDA8 O_3 observations available from each site in 2016 (top). Total number of daily MDA8 O_3 observations in each month of 2016 (bottom).

2.3 O_3 data fusion model

We use multivariate ordinary least squares regression to model the relationship between the individual model components and observed MDA8 O_3. Regression parameters provide estimates of the spatial and temporal model bias attributable to each individual O_3 component. The regression model for ozone mixing ratio O_3 on day *d* and location (*lon*, *lat*, *z*) is formulated as follows:

$$O_3 = \sum_i \alpha_i O_{3_i}^{simulated} + \varepsilon$$

Where:

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

$$\alpha_i = \alpha_{0,i} + \alpha_{x,i} lon + \alpha_{y,i} lat + \alpha_{z,i} z + \alpha_{sin,i} \sin(d) + \alpha_{cos,i} \cos(d)$$

d is day of year in radians

z is elevation above sea level

lon, *lat*, *z*, *sin(d)*, and *cos(d)* are normalized to zero mean and unit standard deviation (Table S6)

$\varepsilon \sim N(0, \sigma^2)$

index *i* represents different sets of O_3 components. Specifically, we consider four sets of *i*:

$i \in \{USA, USB\}$ (PA and EQUATES)

$i \in \{USA, NAT, INTL\}$ (PA only)

$i \in \{USA, NAT, LINTL, CANMEX\}$ (PA only)

$i \in \{USA, USB\_NOSTRAT, STRAT\}$ (EQUATES only)

Each O_3 component is multiplied by the alpha adjustment factor which varies as a function of space and time. The longitude and latitude terms are intended to capture the spatial variability of O_3 biases while the z term is intended to capture biases in O_3 related to elevation. The sinusoidal day of year terms are intended to capture the cyclical nature of O_3 production and to identify any seasonal dependence in O_3 biases. The modeled O_3 components do not add up to observed O_3 because of biases in the model or its inputs. The CMAQ-simulated O_3 components are adjusted by applying estimated regression coefficients to the gridded data so that the sum of the components more closely aligns with observed O_3. The inferred model bias is then calculated as the difference between the original simulated O_3 and adjusted O_3. A more complex method (e.g., nonlinear regression or machine learning) may give a better fit to observed O_3, but the interest here is to estimate potential biases in the modeled O_3 components which is more straightforward with a linear regression. Empirical orthogonal function (EOF) analysis was used to further explore the spatial and temporal structure of the inferred bias fields and is discussed in the SI.

A separate regression model is developed for each model resolution and USB O_3 component split. There are three model resolutions and three USB O_3 splits for the PA simulations, resulting in nine PA models. There are two model resolutions for the EQUATES simulations. The 12 km EQUATES data has two USB O_3 splits while the 108 km EQUATES data has one USB O_3 split, resulting in three EQUATES models. For the PA models, only 2016 data is used since these simulations are for only that year. The

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

models are trained on both 2016 and 2017 data for the EQUATES data. The location and sampling schedule of the monitoring sites overrepresent the eastern US, low elevations, and high O_3 season which may impact how representative the results are for non-monitored locations. Overfitting of the regression model is tested using three cross-validation approaches in which the data are split in both space and time, in space only, and in time only. In the first approach (spatial and temporal withholding), 10% of all observational data are randomly selected and reserved as a test set while the remaining 90% are used as the training set. In the second approach (spatial withholding), data from 10% of randomly selected observation sites are used as a test set while data from the remaining 90% of sites is used as the training set. In the third approach (temporal withholding), data from 10% of randomly selected days of the year are used as a test set while data from the remaining 90% of days of the year are used as the training set. The root mean square error (RMSE) and mean bias for the test and training set are compared to evaluate the potential for the model to overfit the data.

3 Results and discussion

3.1 CTM results

For both the PA and EQUATES simulations, the 12 km simulations have the best performance for MDA8 O_3 as indicated by the normalized mean bias (NMB). The 12 km PA simulations were biased high for 2016 (NMB=1.2%) while the 12 km EQUATES simulations were biased low for 2016 and 2017 (NMB=-3.7% and -5.1%). The 36 km and 108 km PA simulations were biased high over the US for 2016 (NMB=5.2% and 10.0%). The 108 km EQUATES simulations were also biased high over the US for 2016 and 2017 (NMB=2.8% and 0.5%). The two sets of simulations are broadly consistent with one another for BASE, USA, and total USB O_3 which are common to both. Details on the contributions from the different O_3 components in the PA and EQUATES simulations follow.

CMAQ-simulated O_3 from the PA simulations show similar results across the three different model resolutions for USB O_3 sources (Figure 2; Table 2). Simulated USA O_3 tends to increase with coarser model resolution which results in corresponding increases in BASE O_3. This is to be expected with coarser resolutions as more localized NO_x emissions are spread out over larger areas leading to

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

increased simulated O_3 production in NO_x-limited areas and potentially to decreases in NO_x titration in high-NO_x areas. NAT makes the largest contribution to annual average O_3 across the US with a larger contribution in the western US (~55% of BASE) than in the eastern US (~45% of BASE). USA O_3 is the second largest component of annual average O_3 with a larger contribution in the eastern US (~35% of BASE) than in the western US (~20% of BASE). There are a small number of US grid cells with negative annual averages for USA O_3. This means that USB O_3 was greater than BASE O_3 and indicates that anthropogenic emissions suppress O_3 through NO_x titration. LINTL impacts the western US (~15% of BASE) more strongly than the eastern US (~10% of BASE). Both NAT and LINTL tend to be higher at higher elevations, suggesting that some of the effects from NAT and LINTL are from O_3 in the free troposphere. In spring O_3 lifetimes are longer, and trans-Pacific transport of O_3 is more likely which is consistent with the spring peak in LINTL (Liu et al., 1987). The other components and BASE O_3 peak in the summer with some exceptions (Figure 3). In the southeast US, NAT is lower during summer compared to surrounding areas and is lower than NAT in the southeast US during spring. This is likely because O_3 loss through reaction with biogenic VOCs (which peak in the summer and are abundant in the southeast US) reduces O_3 under the extremely low NO_x conditions with zero anthropogenic emissions. The CANMEX contribution to O_3 is small except at some locations along the border with Mexico where the contributions can be high, especially in the summer. For US grid cells within 100 km of the border with Canada, the annual average impact is ~2 ppb while for US grid cells within 100 km of the border with Mexico, the annual average impact is ~5 ppb.

Table 2. Summary of annual average of O_3 components for the Policy Assessment set of simulations. Averages are shown for all of the US and separately for the eastern and western US with a longitude of 97 °W serving as the east-west dividing line. The mean across all grid cells within the given area is shown along with the minimum and maximum for any grid cell within the given area in parentheses.

	BASE	USA	NAT	LINTL	CANMEX
PA 12 km					
all US	39 (18, 56)	10 (-12, 23)	20 (15, 30)	6 (4, 10)	2 (-4, 9)
eastern US	39 (28, 49)	13 (2, 23)	18 (15, 21)	4 (4, 9)	1 (1, 6)
western US	40 (18, 56)	7 (-12, 23)	22 (15, 30)	7 (4, 10)	2 (-4, 9)
PA 36 km					

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

all US	40 (28, 62)	11 (2, 30)	20 (15, 28)	6 (4, 10)	2 (1, 16)
eastern US	40 (28, 55)	14 (4, 28)	18 (15, 21)	4 (4, 9)	1 (1, 5)
western US	40 (30, 62)	8 (2, 30)	22 (15, 28)	7 (4, 10)	2 (1, 16)
PA 108 km					
all US	42 (30, 70)	11 (3, 42)	21 (16, 28)	5 (3, 10)	2 (1, 9)
eastern US	42 (30, 70)	15 (4, 42)	19 (16, 23)	4 (3, 6)	1 (1, 4)
western US	42 (31, 54)	8 (3, 20)	23 (16, 28)	6 (3, 10)	2 (1, 9)

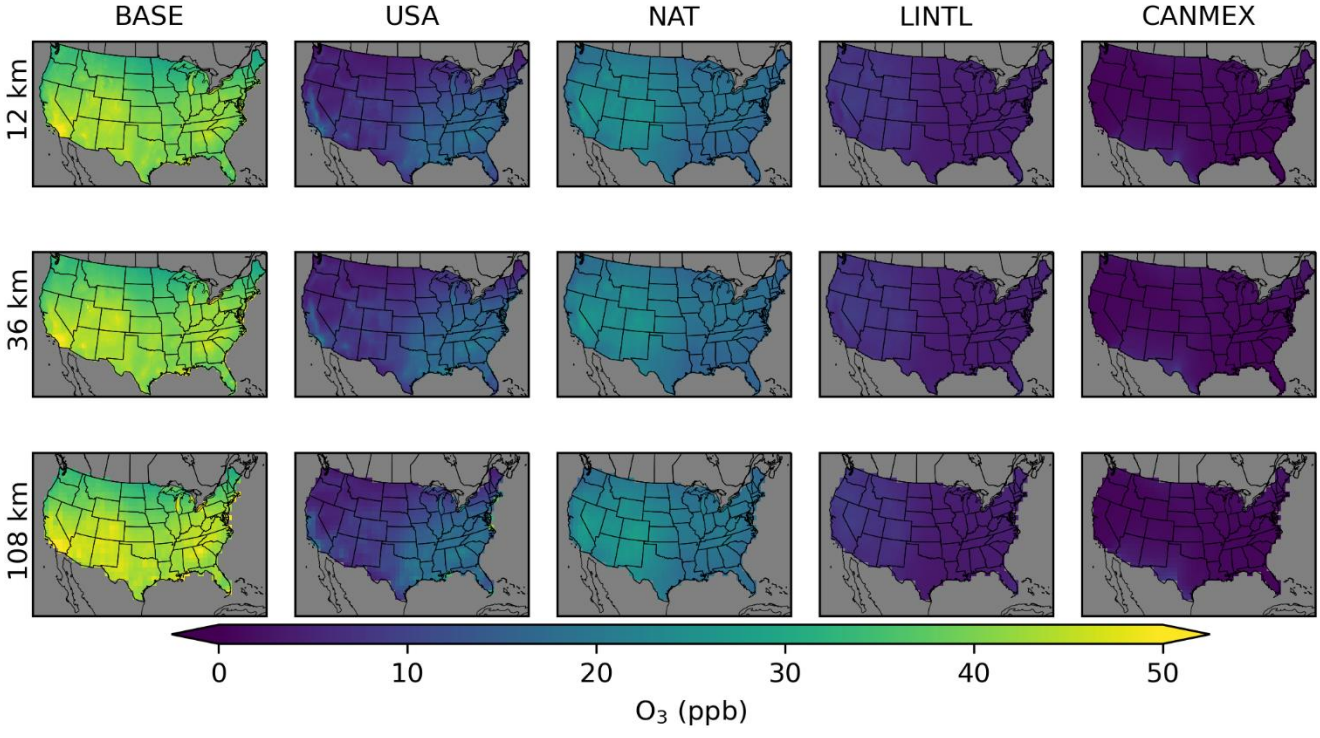


Figure 2. Annual average O_3 from Policy Assessment CMAQ simulations. Results are shown for 12 km (top row), 36 km (middle row), and 108 km (bottom row) horizontal resolutions. O_3 concentrations include total (BASE) O_3 as well as O_3 components from USA, NAT, LINTL, and CANMEX sources.

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

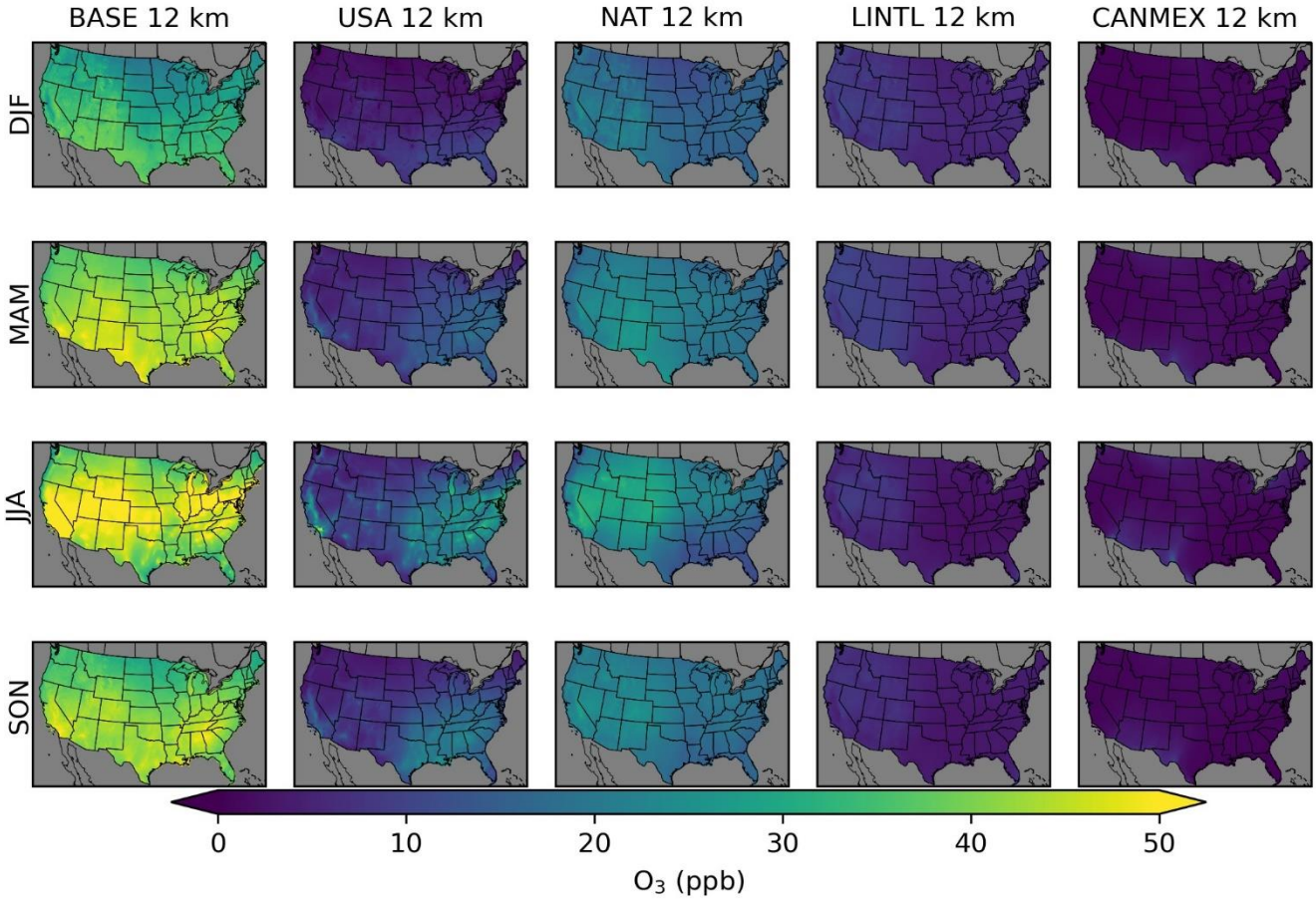


Figure 3. Seasonal average O_3 from Policy Assessment CMAQ simulations. Results are shown for 12 km horizontal resolution for winter (DJF), spring (MAM), summer (JJA), and fall (SON). Seasonal averages for the 36 km and 108 km simulations are provided in the SI (Figures S1 and S2). O_3 concentrations include total (BASE) O_3 as well as O_3 components from USA, NAT, LINTL, and CANMEX sources.

The second set of simulations (EQUATES) split USB O_3 to different components compared to the PA simulations. The use of different USB O_3 components provides additional insight into the source-specific biases in USB O_3. CMAQ simulated O_3 results from the 2016 EQUATES simulations are comparable to the results from the PA simulations for the 12 km simulations, though the EQUATES simulations have slightly less O_3 from USA and more from USB compared to the PA simulations (Figure

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

4; Table 3). USA O_3 contributed ~20% of annual average BASE O_3 across all US model grid cells (~25% for PA simulations). Like in the PA simulations, the contribution to USA O_3 was higher in the eastern US (~25% of BASE) than in the western US (~15% of BASE). STRAT O_3 is higher in the western US, especially at higher elevations, which is consistent with previous studies (Jaffe et al., 2018). On average, STRAT O_3 is 40% of BASE O_3 in the western US and 34% of BASE O_3 in the eastern US. STRAT O_3 represents an upper bound of stratospheric influences because the tracer species used for its calculation in this study does not undergo chemical losses. Non-STRAT O_3 (i.e., USB_NOSTRAT) contributes 47% of annual average BASE O_3 in the western US and 42% in the eastern US. USB_NOSTRAT is likely underestimated in regions and seasons with more active chemistry due to the use of the chemically inert tracer species used to calculate USB_NOSTRAT. The 108 km hemispheric CMAQ (H-CMAQ) results for the EQUATES and PA simulations are similar on average but do have some notable differences. The H-CMAQ simulations are similar in their simulation of USB O_3. The USA O_3 contributions are also similar on average, though the PA simulations have higher maximum values compared to the EQUATES simulations which leads to higher maximum values of BASE O_3.

Table 3. Summary of annual average of O_3 components for the EQUATES set of simulations. Averages are shown for all of the US and separately for the eastern and western US with a longitude of 97 °W serving as the east-west dividing line. The mean across all grid cells within the given area is shown along with the minimum and maximum for any grid cell within the given area in parentheses.

	BASE	USA	USB	USB_NOSTRAT	STRAT
EQUATES 12 km					
all US	39 (22, 51)	7 (-4, 18)	32 (24, 44)	17 (8, 23)	15 (12, 22)
eastern US	38 (30, 45)	9 (1, 15)	29 (24, 36)	16 (8, 23)	13 (12, 19)
western US	40 (22, 51)	5 (-4, 18)	35 (25, 44)	19 (12, 22)	16 (12, 22)
EQUATES 108 km					
all US	41 (31, 49)	8 (2, 18)	33 (26, 41)	---	---
eastern US	40 (31, 49)	10 (3, 18)	30 (26, 38)	---	---
western US	41 (32, 49)	6 (2, 12)	36 (29, 41)	---	---

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

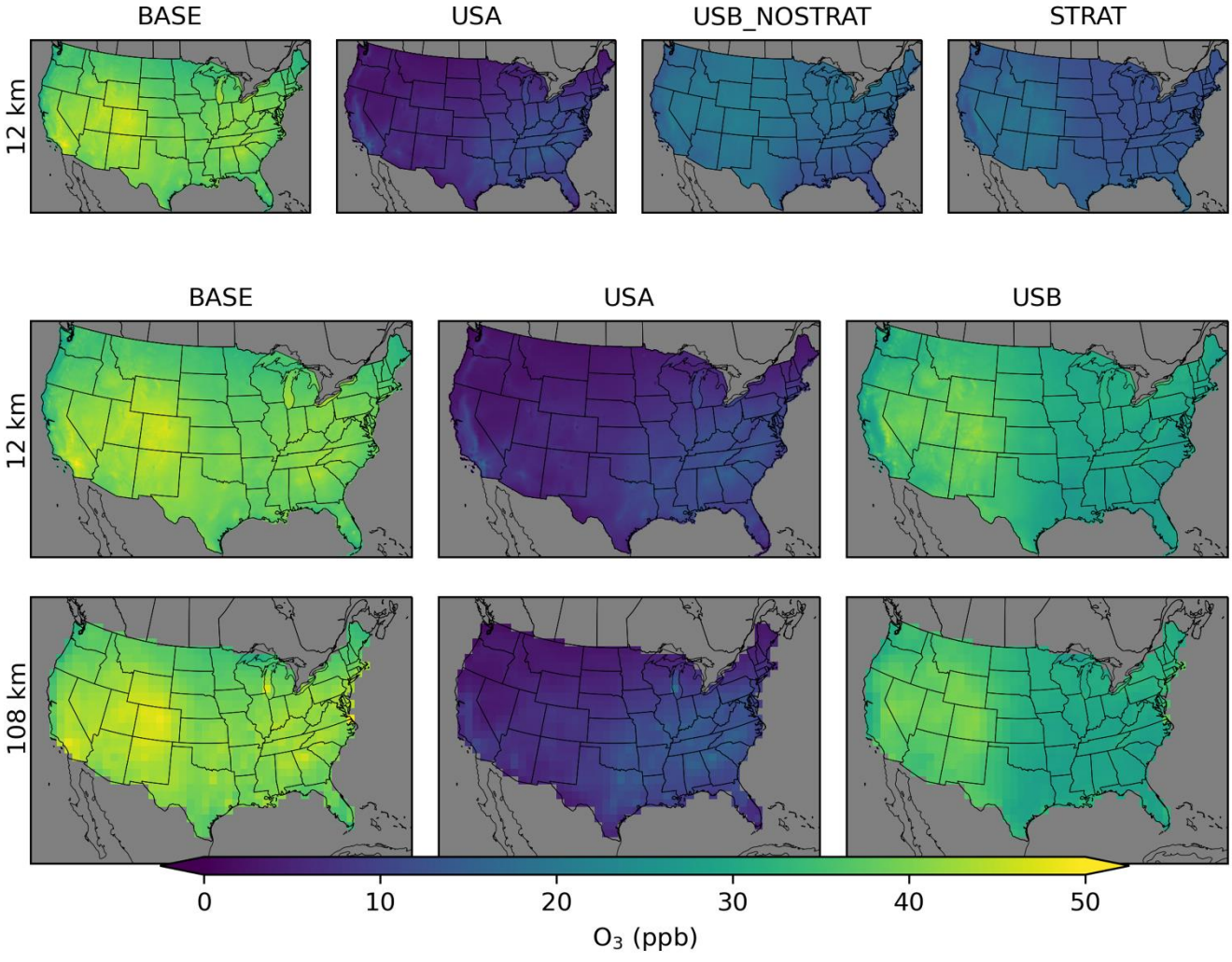


Figure 4. Annual average O_3 from EQUATES CMAQ simulations. Results are shown for 12 km resolution (top and middle rows) and 108 km (bottom row). O_3 concentrations include total (BASE) O_3 as well as O_3 components from USA, USB_NOSTRAT, and STRAT sources for 12 km. For both the 12 km and 108 km simulations, O_3 concentrations of BASE, USA, and total USB are also shown.

BASE O_3 in EQUATES is highest in the summer (Figure 5). USB O_3 is the highest during spring throughout most of the US. In parts of the Mountain West, USB O_3 is highest during the summer (Figures S3 and S4). The STRAT O_3 tracer is the highest in the western US. Much of the western US has STRAT

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

O_3 at about the same level in the spring and summer. In the southeastern US, STRAT O_3 is highest in the summer while in the northeastern US, there are similar levels of STRAT O_3 in the spring and summer. STRAT O_3 is elevated in the summer because of the lack of chemical sinks due to the inert tracer species used to estimate STRAT O_3. Most previous studies have indicated that stratospheric O_3 peaks in the spring (Lin et al., 2015). The stratospheric contribution to O_3 from H-CMAQ calculated using the decoupled direct method (which does account for chemical losses) also showed higher stratospheric contributions in spring than in summer (Mathur et al., 2022). The higher summer STRAT O_3 here is explained by the lack of chemical losses due to the tracer method used. Potential biases are explored further in Section 3.3. USA O_3 is highest in the summer in the eastern US and in California, consistent with the PA simulations. Non-STRAT USB O_3 is relatively uniform outside of summer, though it tends to be slightly lower in the southeast and higher in the western US.

The results from both the PA and EQUATES simulations indicate that USB O_3 contributes more than USA O_3 to BASE O_3 on an annual average basis. Simulated USB O_3 is higher in the western US than in the eastern US due to greater impacts from both natural and non-domestic anthropogenic sources. Simulated USA O_3 is higher in the eastern US than in the western US due to the higher population density and consequently greater anthropogenic emissions. The contributions from USA O_3 peak in the summer which causes BASE O_3 to peak in the summer as well. USB O_3 varies by season but is not as seasonally variable as USA O_3. These results are broadly consistent with previous efforts to quantify USB and USA O_3 using CTMs (Mcdonald-Buller et al., 2011; Jaffe et al., 2018).

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

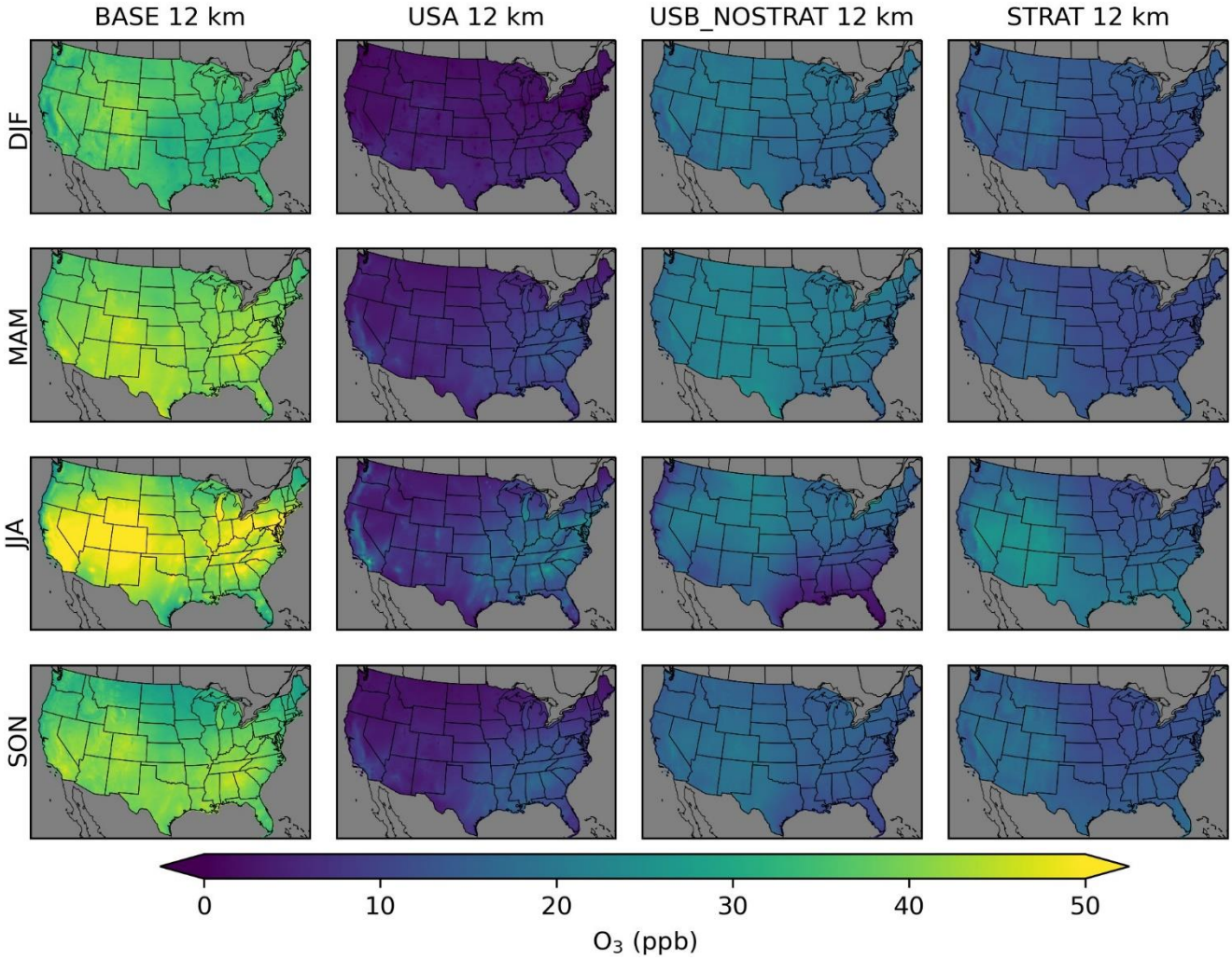


Figure 5. Seasonal average O_3 from EQUATES CMAQ simulations. Results are shown for 12 km horizontal resolution for winter (DJF), spring (MAM), summer (JJA), and fall (SON). O_3 concentrations include total (BASE) O_3 as well as O_3 components from USA, USB_NOSTRAT, and STRAT sources. Seasonal averages for the other USB O_3 split cases are provided in the SI (Figures S3 and S4).

3.2 Cross-validation of regression modeling

Overfitting is tested using a cross-validation analysis as described in Section 2.2. Three different cross-validation methods are used: spatial and temporal withholding, spatial withholding, and temporal

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

withholding. The parameters derived from the training set are then used to predict the observed O_3 in the test set. The RMSE and mean bias with respect to the true observations of both the training and test set are compared to one another (Table 4; Tables S7 and S8). For each of the three cross-validation methods, the RMSE and mean bias of the training and test sets are similar to one another. This indicates that the model is not overfitting and is generalizable to data outside of its training data, providing confidence that we can apply the regression models to the gridded CTM results to estimate the bias in O_3 and individual O_3 components across the US.

Table 4. Summary of performance for cross-validation of O_3 data fusion model. Values shown are the average over all regression model cases. RMSE and mean bias statistics for individual cases are provided in Tables S7 and S8. The performance for the BASE O_3 simulations prior to applying the bias adjustment is also provided for comparison.

metric	BASE simulations	spatial and temporal withholding		spatial withholding		temporal withholding	
		training	test	training	test	training	test
RMSE (ppb)	9.53	7.80	7.83	7.83	7.58	7.81	7.79
mean bias (ppb)	1.13	-0.19	-0.20	-0.19	-0.63	-0.19	0.38

3.3 Inferred CTM biases

The coefficients from the regression models (Tables S9 – S12) are applied to the gridded CTM data to calculate adjusted values of each O_3 component. The inferred CMAQ bias for each component is the difference between the original CMAQ-simulated value and the adjusted value. The inferred bias in BASE O_3 is the original CMAQ-simulated BASE O_3 minus the sum of adjusted O_3 components. For the PA simulations, there is a residual anthropogenic component of BASE O_3 that is not apportioned to either USA or INTL sources due to the effects of non-linear chemistry (Table S2). The residual anthropogenic component is equal to BASE – NAT – INTL – USA. This means that the sum of biases in the individual components do not add up to the bias in BASE O_3 as the residual anthropogenic component was not included in the adjusted O_3 results. In the PA simulations, BASE O_3 is inferred to be biased high in most of the Eastern US as well as in California and Arizona (Figure 6). USA O_3 is inferred to be biased high in the same areas. Reducing the amount of USA O_3 improves the fit to BASE O_3 which is suggestive that

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

biases in the effects from US anthropogenic emissions contribute to the high biases inferred in BASE O_3. The inferred high biases in BASE and USA O_3 increase with increasing coarseness of model resolution. Similarly, the high bias increases with coarser model resolution in the CANMEX component along the border with Mexico.

There are offsetting inferred biases in the LINTL and NAT O_3 components in much of the western US. The offsetting inferred biases may reflect an inability of the regression model to separate the signals from LINTL and stratospheric O_3. LINTL and stratospheric O_3 are expected to impact sites at similar spatial and temporal scales, with larger impacts expected at high elevations in the western US during spring. Stratospheric O_3 effects are not limited to episodic intrusion events but also come from constant entrainment of stratospheric air to the free troposphere. The impacts from LINTL are primarily from long-range transport in the free troposphere, so stratospheric O_3 and LINTL are expected to be correlated. The regression model may be assigning bias due to stratospheric O_3 to LINTL because the CTM-modeled LINTL component has more correlation with the stratospheric O_3 impact than the CTM-modeled NAT component. This could result in the regression model adjusting LINTL upwards (i.e., inferred negative bias) to add stratospheric O_3. The NAT O_3 is then adjusted downwards (i.e., inferred positive bias) in the same locations because some of the effects of stratospheric O_3 are captured in the CTM-modeled NAT component but need to be offset because of the O_3 that was added to the LINTL component. This indicates a limitation of this method in that it is sensitive to correlation between modeled O_3 components. Correlation of the O_3 components is a major confounding issue in this analysis. In interpreting the results, it is necessary to consider both the inferred biases and the correlation of the components together.

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

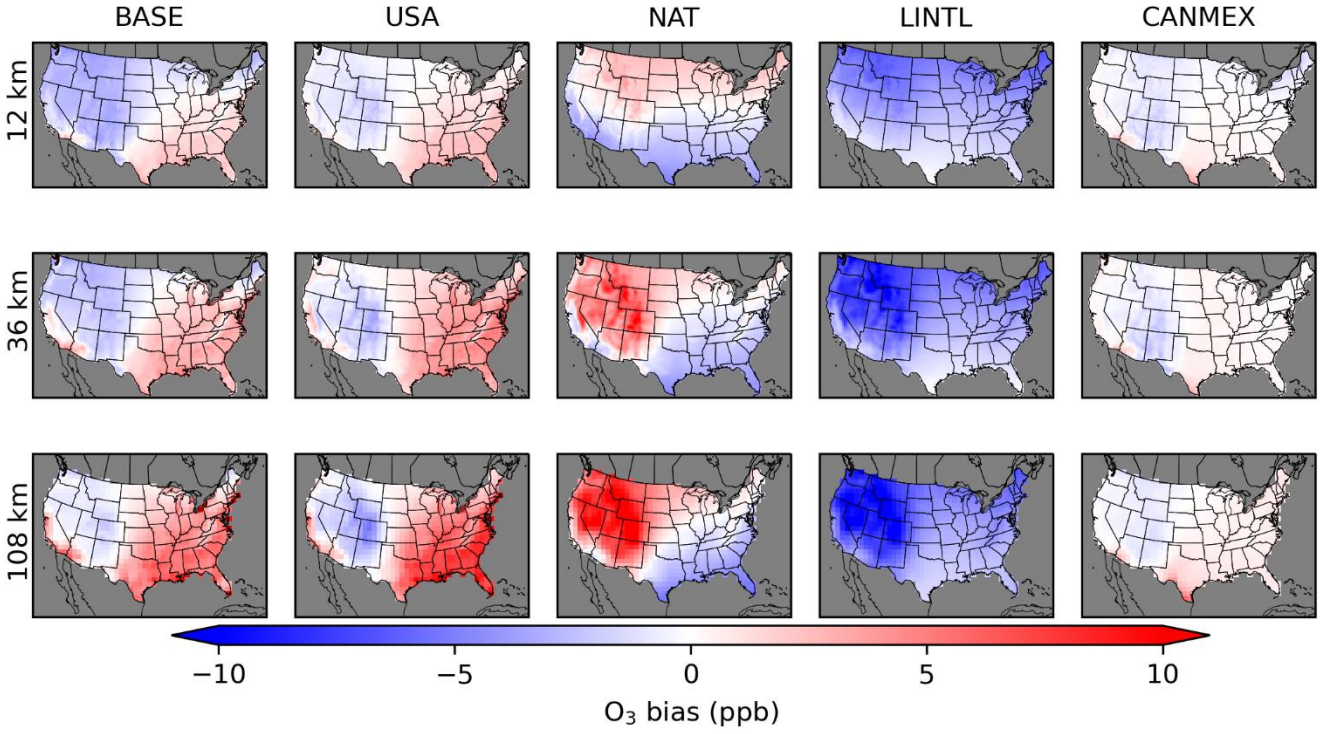
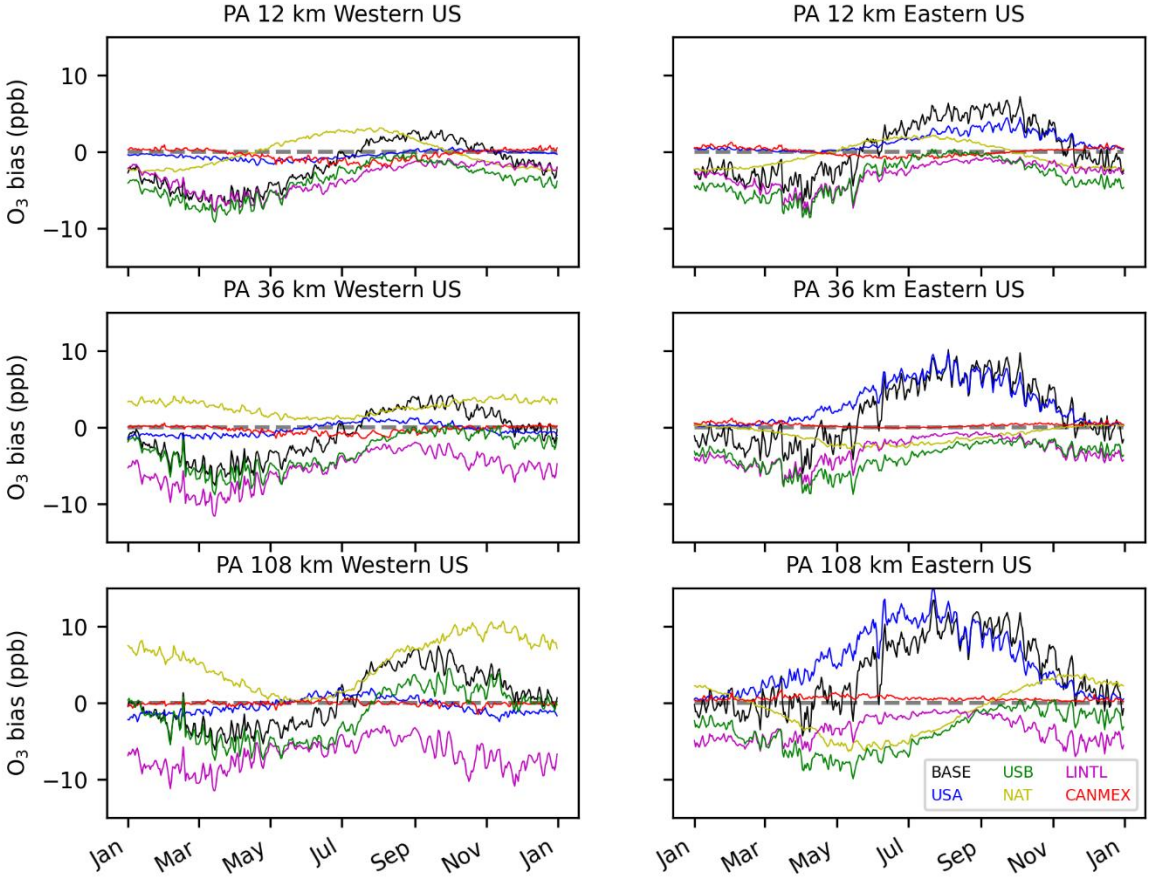


Figure 6. Annual average of inferred O_3 model bias from Policy Assessment CMAQ simulations. Results are shown for 12 km (top row), 36 km (middle row), and 108 km (bottom row) horizontal resolutions. O_3 concentrations include total (BASE) O_3 as well as O_3 components from USA, NAT, LINTL, and CANMEX sources.

In the temporal trends of inferred BASE O_3 bias, the PA simulations show a consistent low bias in winter and spring and high bias in summer and fall which is consistent across model resolution scales (Figure 7). There is also a consistent high bias in USA O_3 in summer and fall in the eastern US which increases with coarser model resolution. Inferred bias in USA O_3 in the western US has some small seasonal variability but is near zero on average. The seasonal patterns of LINTL bias have the largest underestimate in the winter and spring and the smallest underestimate in late summer and early fall. The temporal trend of NAT differs in the 12 km simulation compared to the 36 km and 108 km simulations. In the 12 km simulation, NAT biases are higher in the middle of the year than in the beginning and end of the year. In the 36 km and 108 km simulations, the opposite is found. This change in sign is a result of

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

changes in the spatial patterns of NAT inferred bias in different seasons. In the 12 km simulation, NAT is inferred to be biased low in the southern part of the US and biased high in the northern part of the US. In the 36 km and 108 km simulations NAT is inferred to be biased low in the eastern US and mostly biased high in the western US, particularly in the Mountain West region. These spatial changes in the seasonal average NAT O_3 bias are enough to change the sign of the US average temporal bias trend. As described before, the offsetting negative LINTL bias and positive NAT bias in the high elevation areas of the western US are thought to be a result of the regression model allocating stratospheric O_3 bias to the LINTL signal while removing some stratospheric O_3 from the NAT signal. CANMEX O_3 biases are very small when averaged across the US since this source primarily affects border areas and only has small impacts elsewhere.



https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

Figure 7. Daily average of inferred O_3 model bias from Policy Assessment CMAQ simulations averaged across US model grid cells in the eastern and western US. A longitude of 97 °W is used as the dividing line between east and west. PA O_3 concentrations include total (BASE) O_3 as well as O_3 components from USA, NAT, LINTL, and CANMEX sources. USB indicates the sum of biases for individual USB components.

The spatial results for the EQUATES 12 km simulations are shown for two O_3 split cases. One case splits USB O_3 to STRAT and non-STRAT sources while the other considers all USB O_3 together. Results show a mostly low bias inferred in BASE O_3 throughout most of the US for the 12 km simulation (Figure 8). For the 108 km H-CMAQ simulation there is a high bias in the eastern US and a low bias in the western US for BASE O_3. Like the PA results there is a high bias in USA O_3 in the eastern US that increases with coarser model resolution. The inferred low bias in the STRAT O_3 component indicates that there is too little stratospheric O_3 in the western US. There is an inferred high bias in STRAT O_3 in the eastern US. The STRAT O_3 results should be interpreted with some caution because the STRAT component comes from a chemically inert tracer. The STRAT O_3 biases are partly offset by opposite biases in the non-STRAT USB O_3. The low biases in STRAT O_3 and the lack of low biases in the non-STRAT USB O_3 gives more evidence to the proposition that the low biases in the LINTL O_3 from the PA simulations is a result of low biases in stratospheric O_3.

In the case where USB O_3 is not split into STRAT and non-STRAT components, the 12 km and 108 km simulations both have low biases in USB O_3, but the magnitude of bias is greater in the 12 km simulation than in the 108 km simulation. This may be a result of differences in the impacts of stratospheric O_3 at the surface level in the H-CMAQ simulation compared to the continental-scale simulation. Differences in the estimation of stratospheric O_3 impacts may arise from differences in how the vertical structure of the model in the H-CMAQ simulations is configured compared to the continental simulations. The UTLS PV O_3 scaling is turned on during the H-CMAQ simulation. For the continental simulation, PV O_3 scaling is turned off because the continental model configuration uses fewer vertical layers and a coarser vertical resolution in the UTLS compared to the H-CMAQ simulations. The stratospheric O_3 influences in the continental simulation are only those influences that are inherited from the lateral boundary conditions. Previous work indicates that O_3 in the upper layers of the continental-

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

scale model is driven mostly by horizontal advection of the lateral boundary conditions (Hogrefe et al., 2018), meaning that if stratospheric intrusion events are captured by the hemispheric-scale simulation, the effects of these events are also expected to be captured by the continental-scale simulation. However, a sensitivity test with UTLS PV O_3 scaling turned on during the continental simulation may be an area for future study. This would require the addition of more vertical layers with finer resolution in the UTLS in the continental simulation to support the PV O_3 scaling parameterization. The differences in vertical structure of the hemispheric and continental simulations can affect the vertical mixing of stratospheric O_3 from upper layers down to the surface which may explain the differences in inferred bias of USB O_3. Alternatively, the differences in USB O_3 biases could also occur due to differences in O_3 production from local USB O_3 sources across model resolution scales and may not necessarily be affected by differences in stratospheric O_3.

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

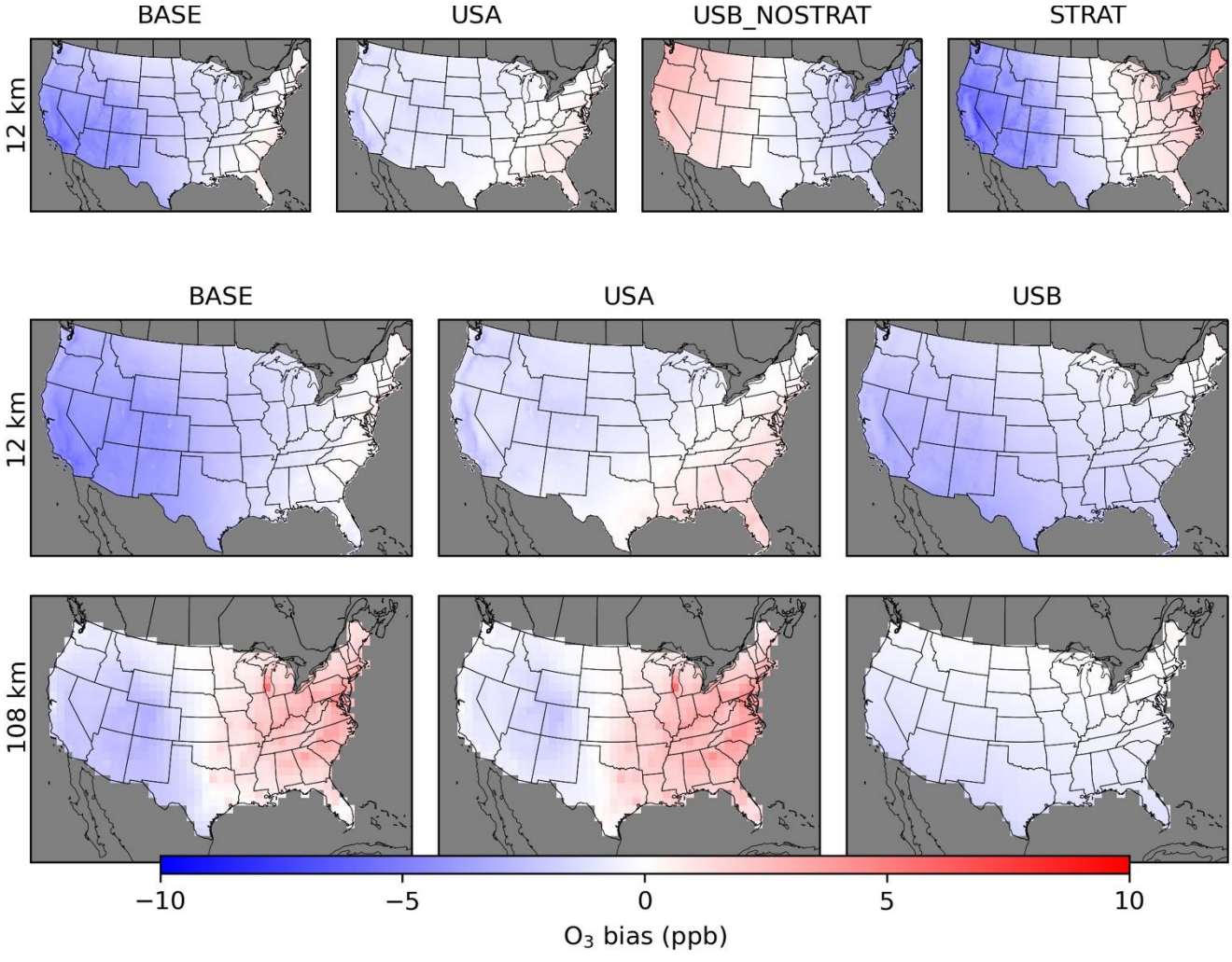
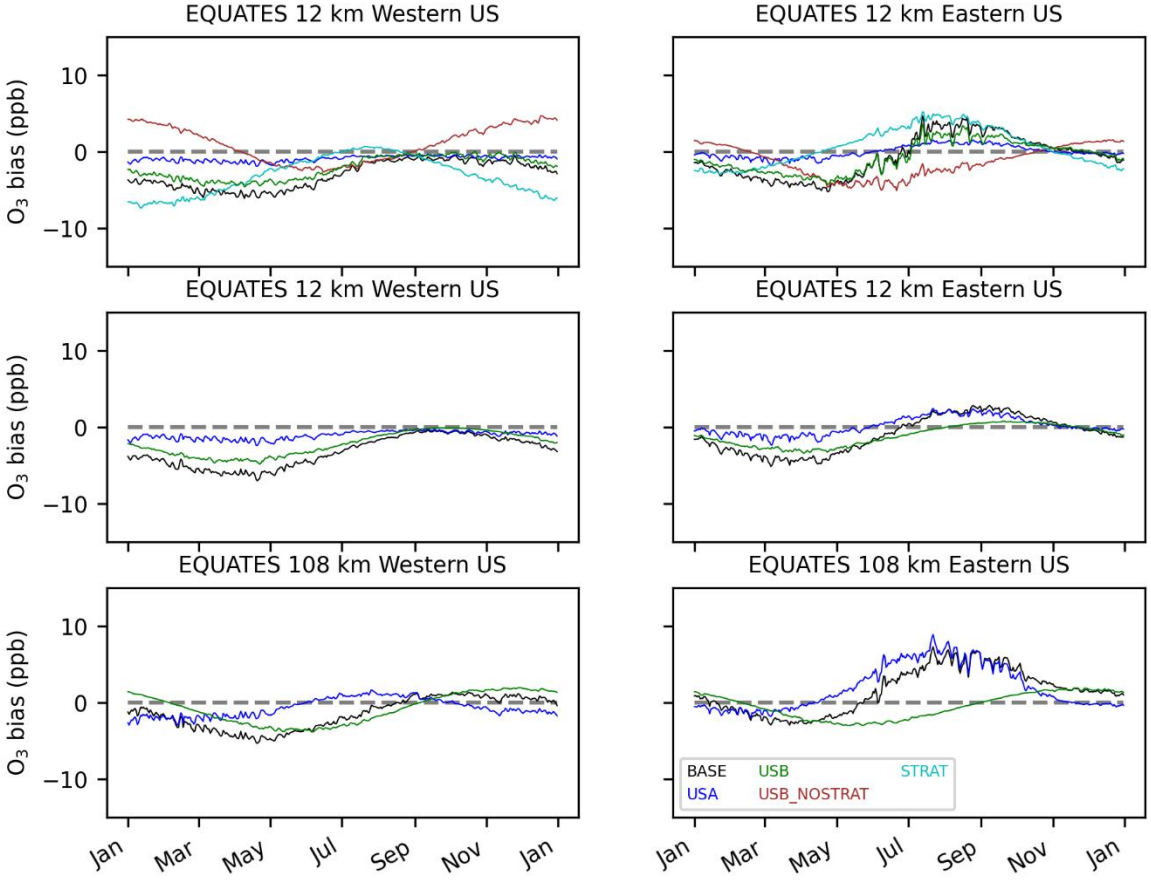


Figure 8. Annual average of inferred O_3 model bias from EQUATES CMAQ simulations. Results are shown for 12 km resolution (top and middle rows) and 108 km (bottom row). O_3 concentrations include total (BASE) O_3 as well as O_3 components from USA, USB_NOSTRAT, and STRAT sources for 12 km. For both the 12 km and 108 km simulations, O_3 concentrations of BASE, USA, and total USB are also shown.

For the EQUATES temporal results, BASE O_3 is biased low in the spring and high in the summer in the eastern US (Figure 9). In the western US, BASE O_3 is biased low throughout most of the year. Averaged across the US, bias is near zero in the summer and fall in the 12 km simulation with high biases in the 108 km simulation during the same period (+1 ppb in summer; +2 ppb in fall). The high biases in

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

BASE O_3 in the eastern US are mostly due to high biases in the USA O_3 component which peak in the summer and continue to be biased high into the fall. The STRAT O_3 component is inferred to be biased low except in the summer and early fall. In the western US, STRAT O_3 bias in the summer is near zero in the summer and fall while in the eastern US, STRAT O_3 is biased high in the summer and fall. The lowest biases in STRAT O_3 occur in the winter. The STRAT O_3 biases are partially offset by opposing biases in the non-STRAT USB O_3. The regression model formulation without the separate STRAT O_3 indicates that there is a low bias in USB O_3 throughout most of the year in the 12 km simulation which is at its lowest in the spring. The 108 km simulations show a low bias for USB O_3 in the spring and summer and high bias in the fall and winter.



https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

Figure 9. Daily average of inferred O_3 model bias from EQUATES CMAQ simulations averaged across US model grid cells in the eastern and western US. A longitude of 97 °W is used as the dividing line between east and west. EQUATES O_3 concentrations include BASE O_3 as well as O_3 components from USA, USB_NOSTRAT, and STRAT sources for 12 km. For both the 12 km and 108 km simulations, O_3 concentrations of BASE, USA, and total USB are also shown. For the case with multiple USB O_3 components, USB indicates the sum of biases for individual USB components.

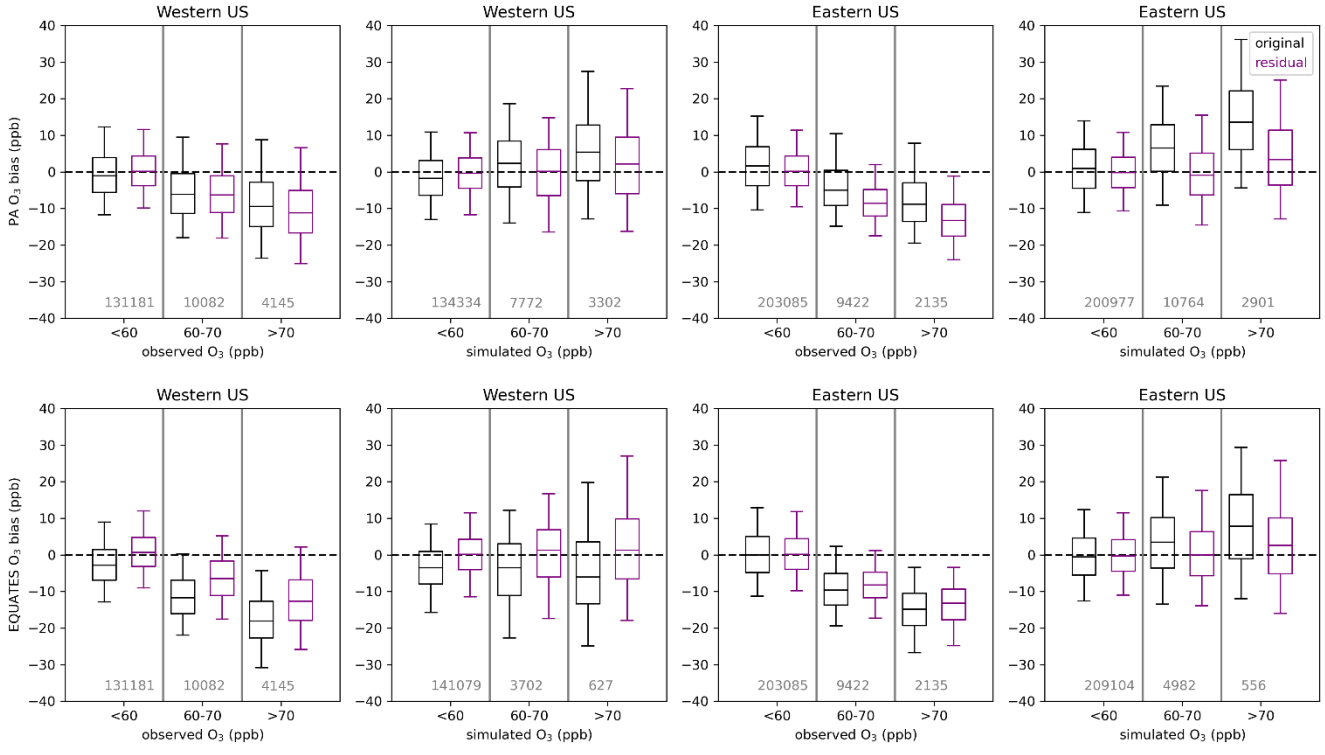
3.4 CTM biases by O_3 concentration

Situations where O_3 exceeds the NAAQS, which is currently set at a level of 70 ppb, are of particular interest. We analyze the different O_3 components at O_3 monitoring sites under cases when O_3 is less than 60 ppb, between 60 and 70 ppb (inclusive), and greater than 70 ppb. These concentration bins are selected because they reflect the current level of the standard (70 ppb) as well as a potential range which might be considered as the level of the standard in the future (60-70 ppb). We compare the results of the analysis when using both simulated and observed O_3 bins. Simulated O_3 has a positive bias on average when simulated O_3 is high and has a negative bias on average when observed O_3 is high, so selection bias influences these results. For this analysis, we consider the 12 km resolution simulations for the PA and EQUATES simulations. The 12 km simulations are the resolution that is typical for simulations that support regulatory analyses. Monitoring sites are split into western or eastern US using a longitude of 97 °W as the dividing line. The division to western and eastern US is done because there are differences in the contribution of US anthropogenic vs. background contributions in the two parts of the country.

The impacts of the linear regression adjustment technique at the observation sites are examined by comparing the original simulated bias to the residual bias (i.e., the sum of the adjusted individual O_3 components minus observed O_3) (Figure 10). The change in bias from the original to residual bias is the inferred bias that has been referenced elsewhere. In all cases when O_3 is binned by simulated O_3 levels, the adjustment brings the bias closer to zero. In the eastern US, high biases at higher simulated O_3 levels were reduced for both the PA and EQUATES simulations. In the western US, low biases when simulated O_3 was below 60 ppb were brought closer to zero for both the PA and EQUATES simulations. At higher

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

simulated O_3 levels, the PA simulations originally had high biases in the western US which were reduced in the adjusted results while the EQUATES simulations originally had low biases in the western US which were improved in the adjusted results. The effects on bias when binning by observed O_3 are mixed. In both the western and eastern US for both the PA and EQUATES simulations, the simulations were originally biased low at higher observed O_3 levels, with the EQUATES simulations more biased low than the PA simulations. The low bias is improved in the EQUATES simulations, but in the PA simulations the bias is either about the same or becomes more biased low. The inability of the adjustment to improve the bias across the range of both observed and simulated O_3 levels is a limitation of this technique. The fitting of multi-axis (lat, lon, season) linear correction factors (α_i) will be strongly influenced by the larger population of lower (O_3 < 70 ppb) concentrations and will only correct the upper end if the bias structure is consistent.



https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

Figure 10. Bias compared to O_3 observations of original simulations (black) and residual bias (purple) obtained as the difference between adjusted O_3 and observations for PA (top row) and EQUATES (bottom row) simulations. The horizontal line shows the median; the box shows the 25th-75th percentiles; the whiskers show the 5th and 95th percentiles. Grey vertical lines separate the boxplots for each O_3 concentration bin. The numbers at the bottom of each panel are the number of data points falling within each concentration bin.

For the PA simulations, the contribution from USA O_3 tends to increase with higher simulated O_3 and with higher observed O_3 (Figure 11), indicating that domestic anthropogenic pollution is driving the highest O_3 concentrations. The contribution from USA O_3 is higher at eastern US sites than at western US sites due to higher anthropogenic precursor emissions in the east. There may also be impacts on USA O_3 in the eastern US from O_3 or precursor pollutants transported from the western to eastern US. The median USA O_3 contribution is biased high (+1 ppb in the western US; +4 ppb in the eastern US) when BASE O_3 is between 60 and 70 ppb with higher median biases (+2 ppb in the western US; +6 ppb in the eastern US) when BASE O_3 exceeds 70 ppb. When observed O_3 is between 60 and 70 ppb, the median USA O_3 contribution is biased slightly low in the western US (-0.2 ppb) and biased high in the eastern US (+2 ppb). Bias is higher in the western US when observed O_3 exceeds 70 ppb (+1 ppb) but is about the same in the eastern US (+2 ppb). Inferred biases of USA O_3 are higher across the range of simulated and observed O_3 levels in the eastern US compared to the western US.

In the western US, NAT O_3 tends to be higher when either simulated or observed O_3 is greater than 60 ppb; however, the distribution of NAT O_3 when O_3 is above 70 ppb is similar to the distribution of NAT O_3 when O_3 is between 60 and 70 ppb. In the eastern US, the distribution of NAT O_3 is similar across the range of simulated and observed O_3 concentration bins but is slightly higher when O_3 is greater than 60 ppb. LINTL makes a small contribution to O_3 across concentration bins and tends to be lower as simulated or observed O_3 increases. CANMEX O_3 is typically very small and only makes significant contributions at a few near-border sites (not shown). The NAT and LINTL O_3 components are biased slightly low at monitoring sites in the western US. For western US sites, the sum of the median biases in USA and USB (i.e., NAT+LINTL+CANMEX) O_3 at monitoring sites is negative across the simulated and observed O_3 concentration bins but gets closer to zero at higher O_3 levels. For eastern US sites, the bias in USA O_3 is predicted to be the main contributor to biases at high simulated O_3 when simulated O_3

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

concentrations exceed 60 ppb. When the O_3 components are binned by observed O_3 rather than simulated O_3, the sum of the median biases in USA and USB O_3 at monitoring sites in the eastern US is negative across the range of simulated O_3 with USB O_3 becoming less negatively biased as observed O_3 increases and USA O_3 becoming more positively biased as observed O_3 increases.

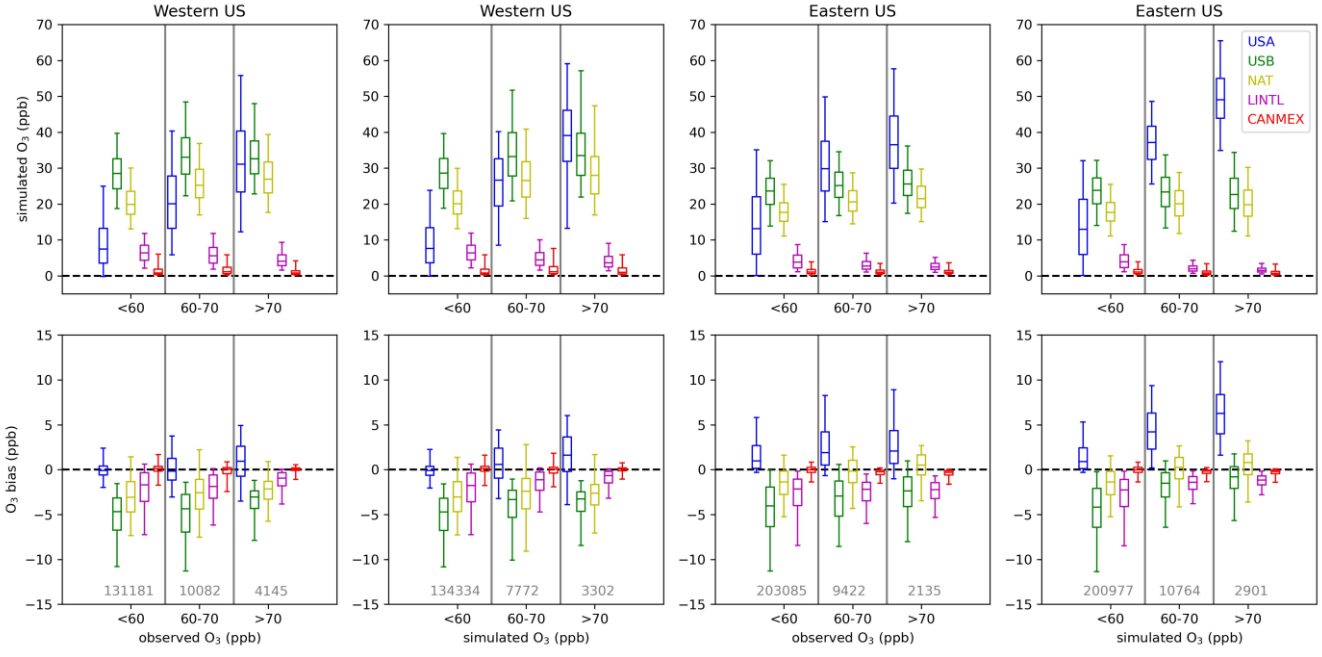


Figure 11. Contributions from the PA simulation (top row) and inferred biases (bottom row) of USA, NAT, LINTL, and CANMEX separated by both observed and simulated BASE O_3 concentration at O_3 monitoring sites. The sum of NAT, LINTL, and CANMEX is shown as USB. The horizontal line shows the median; the box shows the 25th-75th percentiles; the whiskers show the 5th and 95th percentiles. Grey vertical lines separate the boxplots for each O_3 concentration bin. The numbers in the bottom row of panels are the number of data points falling within each concentration bin.

For the 12 km EQUATES simulations, the USA O_3 contribution is similar to the 12 km PA results across the simulated O_3 concentration bins (Figure 12). At higher observed O_3, the EQUATES simulations generally simulate lower USA O_3 compared to the PA simulations. Like in the PA simulations, the USA O_3 contribution increases with increasing simulated and observed O_3, meaning that domestic

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

anthropogenic emissions are mostly driving the highest O_3 levels. There is an inferred negative bias in USA O_3 in the western US which becomes increasingly more negative as simulated or observed O_3 increases. In the eastern US, there is an inferred positive bias in USA O_3 which becomes larger at higher simulated O_3 concentrations (median bias of +0.05, +2, +4 ppb at <60, 60-70, and >70 ppb simulated O_3). There is also an inferred high bias across the range of observed O_3; however, the magnitude is smaller, and the bias does not increase much at higher levels of observed O_3 (median bias of +0.05, +0.5, and +0.6 ppb at <60, 60-70, and >70 ppb observed O_3).

The contribution from STRAT O_3 is higher in the western US than in the eastern US across simulated and observed O_3 concentrations. In the western US, STRAT tends to be higher when either observed or simulated is above 60 ppb. In the eastern US, STRAT O_3 is at similar levels across the range of simulated and observed O_3. In the western US, STRAT O_3 has a negative bias which gets closer to zero when simulated and observed O_3 is above 60 ppb. In the eastern US, STRAT O_3 has a positive bias which gets higher when simulated and observed O_3 are above 60 ppb. In both the western and eastern US, USB_NOSTRAT makes similar contributions across different O_3 concentrations. In the western US, USB_NOSTRAT has a negative bias when simulated or observed O_3 is below 60 ppb and a positive bias when O_3 is above 60 ppb. In the eastern US, USB_NOSTRAT has a negative bias across the range of simulated and observed O_3. The magnitude of the negative bias is smaller when simulated or observed O_3 is below 60 ppb than when O_3 is above 60 ppb.

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

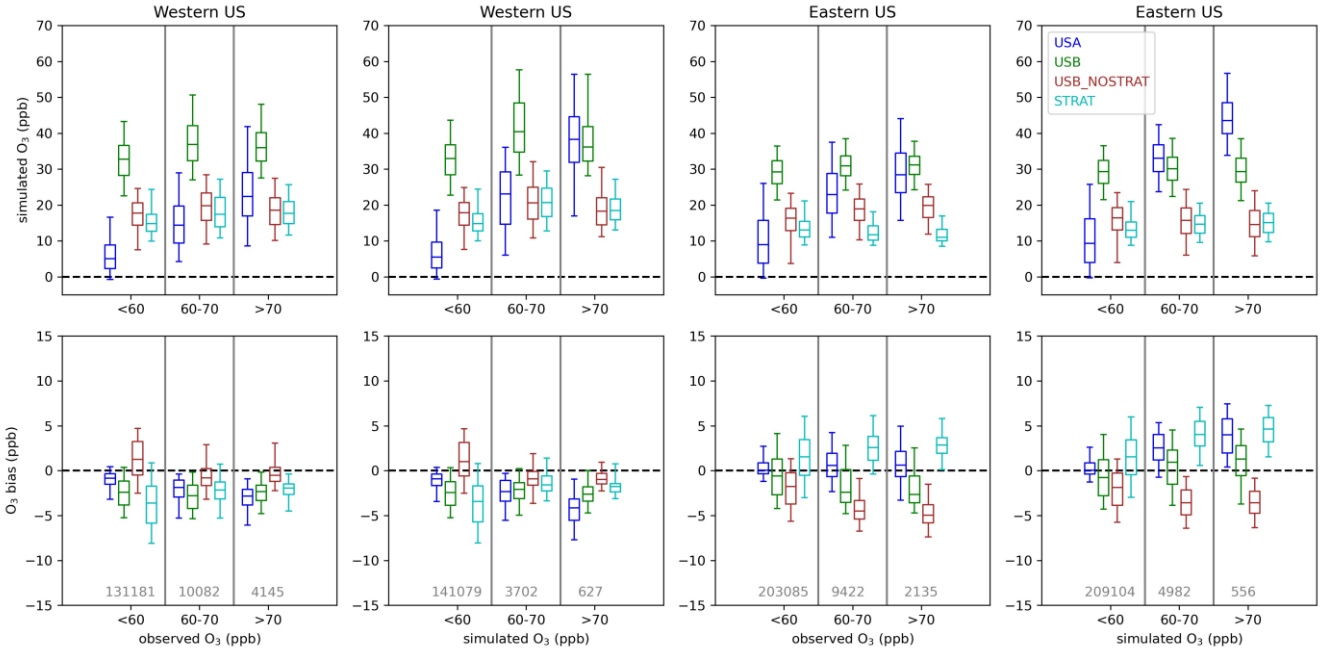


Figure 12. Contributions by the EQUATES simulation (top row) and inferred biases (bottom row) of USA, USB_NOSTRAT, and STRAT separated by both observed and simulated BASE O_3 concentration at O_3 monitoring sites. The sum of USB_NOSTRAT and STRAT is shown as USB. The line shows the median; the box shows the 25th-75th percentiles; the whiskers show the 5th and 95th percentiles. Grey vertical lines separate the boxplots for each O_3 concentration bin. The numbers in the bottom row of panels are the number of data points falling within each concentration bin.

Binning the O_3 contributions and inferred biases by observed and simulated O_3 results in different numbers of data points in each sample. In the western US, there were 4145 instances when observed O_3 exceeded 70 ppb, while there were 3302 (PA) and 627 (EQUATES) instances when simulated O_3 exceeded 70 ppb at a monitoring site, with a large fraction of the observed and simulated exceedances occurring in California. In the eastern US there were 2135 instances when observed O_3 exceeded 70 ppb with 2901 (PA) and 556 (EQUATES) instances when simulated O_3 exceeded 70 ppb. The PA simulations more accurately simulated the number of exceedances compared to EQUATES, though this does not consider the timing or location of exceedances. Given the different number of samples in the observed vs. simulated bins and the lower number of data points for EQUATES simulated O_3 exceeding 70 ppb, it

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

is possible that the population of data points are when simulated O_3 exceeds 70 ppb are not spatially representative of the population when observed O_3 exceeds 70 ppb.

For the western US, the PA simulations largely capture the spatial distribution of exceedances seen in the observations, although the number of exceedances is underestimated (Figure 13). The exceedances from the EQUATES simulations are not very representative of the spatial distribution of observed exceedances in the western US as there are very few sites with more than one or two exceedances outside of California. In particular, the number of exceedances in the Denver, Colorado; Phoenix, Arizona; Las Vegas, Nevada; and Boise, Idaho; areas are underestimated in EQUATES relative to both the PA simulations and observations. Both the PA and EQUATES simulations underestimate the number of exceedances in the state of Utah. For the eastern US, the PA simulations generally capture the spatial distribution of observed exceedances but simulate too many exceedances. This is particularly notable in the northeastern US and along the Gulf Coast. The EQUATES simulations underestimate the number of exceedances, although the spatial distribution is generally similar to the observations. The degree of spatial representativeness provides additional context for interpreting the findings for the O_3 component contributions and biases binned by O_3 levels. For the western US, the findings for instances when O_3 exceeds 70 ppb are not applicable to the western US more broadly. There are a limited number of instances when O_3 exceeds 70 ppb in the western US outside of California. These results are mostly indicative of conditions in the Los Angeles area and in the Central Valley in California. This applies especially to the EQUATES results, but it is also the case for the PA simulations and the observations. For the eastern US, on the other hand, there is enough spatial variability in the observations as well as both sets of simulations to interpret the findings for the eastern US more generally. These results are informative in an average sense but are not expected to hold in all cases when applied to specific monitoring sites or to specific days (e.g., fourth highest O_3). The biases for bins 60-70 ppb and greater than 70 ppb should be interpreted with caution because the inferred biases apply the mean tendency to these high concentration subpopulations.

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

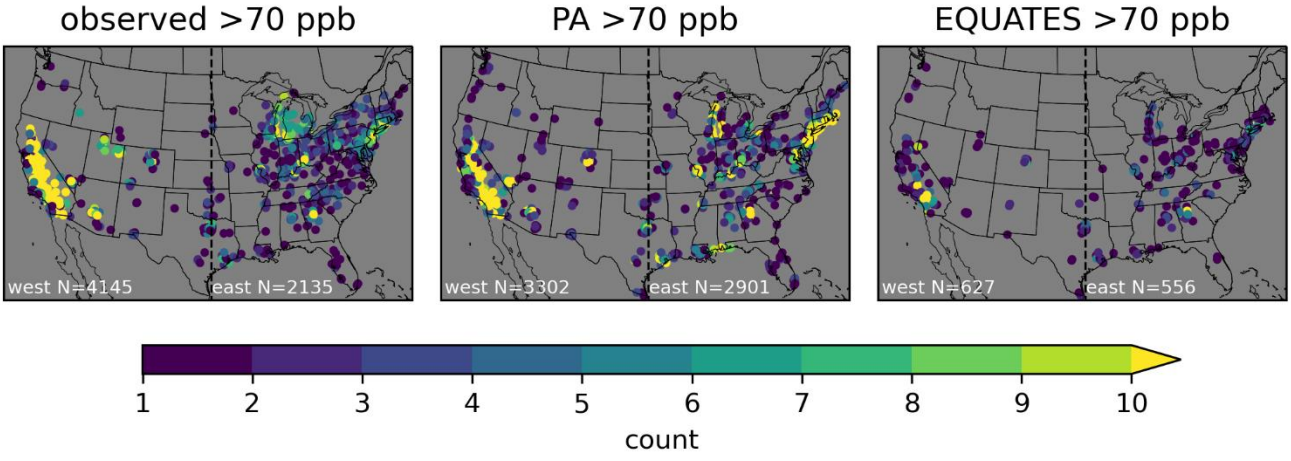


Figure 13. Spatial distribution of the number of times O_3 exceeded 70 ppb for observed and simulated O_3. The circles show the locations of sites, and the color indicates the number of times O_3 exceeds 70 ppb at each site for observations (left), PA simulation (middle), and EQUATES simulation (right). Only sites with at least one exceedance are shown. The black dotted line shows the longitude of 97° W which is used to divide west and east.

4 Conclusions

In this work, we use two sets of CMAQ simulations to analyze the contributions to USB O_3 from different sources. Naturally occurring sources, long-range international anthropogenic pollution, and short-range international anthropogenic pollution from Canada and Mexico are separately considered for one set of simulations. In the other set of simulations, stratospheric and non-stratospheric sources of USB O_3 are also considered separately. We also consider the contribution to total O_3 from US domestic anthropogenic sources. The measurement-model data fusion approach for apportioning bias to USA and USB O_3 components from our previous study (Skipper et al., 2021) was extended to identify biases in separate USB O_3 components. The results generally confirm previous high-level results, but provide new insights from additional components and more detailed analysis.

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

Results indicated that USA O_3 was consistently inferred to be biased high in the eastern US where domestic anthropogenic emissions are the dominant contributor to total O_3, with increasingly higher biases with coarser model resolution and at higher simulated O_3 concentrations. This is consistent with our previous findings. This does not necessarily imply that the trend of decreasing biases with finer resolutions would continue at resolutions finer than 12 km as we have not tested this approach at those resolutions. The finding that USA O_3 biases increase with higher O_3 does not hold when O_3 is binned by observed rather than simulated concentrations. There is much less variation in the USA O_3 bias across the range of observed O_3 than for simulated O_3. Although the choice of binning O_3 by observed or simulated levels changes the sample of data, the results for the eastern US are generalizable to this part of the country because the samples have consistent spatial representation across the eastern US. In the western US, USA O_3 was inferred to be biased high at higher O_3 levels for the PA simulations and biased low at higher O_3 levels for the EQUATES simulations. These differences are explained by the use of different emission inventories in the two sets of simulations. Regardless, the findings for inferred O_3 biases at higher O_3 levels in the western US are not broadly applicable to the entire western US because the sample that these findings are based on is dominated by sites in California. There are relatively few sites in other states in the western US that contribute to this sample, so the results are not likely to be indicative of conditions in other parts of the western US. The correction of USB components provided consistent results with previous studies, but more detail. Like Skipper et al. (2021) and Hosseinpour et al. (2024), simulated USB O_3 was inferred to be biased slightly low overall. The original simulated annual averages of USB O_3 across all the PA and EQUATES modeling configurations considered here ranged from 30-33 ppb while the adjusted annual average USB O_3 ranged from 31-34 ppb. This work separated USB into natural, short-range international, and long-range international and each had distinct seasonality to the inferred bias. Short-range international was marginally high-biased in spring/winter and marginally low-biased in summer. The contribution from natural and long-range international have larger seasonality, which are slightly out of phase. Natural bias was low in winter, but high in summer peaking in July. Long-range international was consistently low-biased with a minimum in April and a maximum (near unbiased) in August-September. From May to October, the natural and long-range international biases are largely offsetting while they are reinforcing in other parts of the year.

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

The seasonality of inferred long-range international bias highlights a key uncertainty in correlative bias attribution. The biases associated with long-range international may be misattributed due to the difficulty of the regression model formulation to isolate stratospheric influences from other natural sources. Stratospheric O_3 is expected to have similar temporal and spatial patterns to LINTL, with contributions being higher in spring and at high elevations. It is suspected that the regression model formulation may be assigning a negative bias in LINTL to make up for missing stratospheric O_3 that has a similar pattern to LINTL while at the same time assigning a high bias for NAT to reallocate some of stratospheric O_3 that is present in NAT to LINTL instead. Results for the STRAT O_3 tracer in the second set of simulations support the idea that there is missing stratospheric O_3 at the surface level in the western US as the STRAT O_3 is inferred to be biased low. Taken together, there is an overall low bias in the simulated USB O_3 that is most pronounced in the spring. This may be a result of too little stratospheric O_3 reaching the surface. Photolysis of particulate nitrate over oceans has been found to increase O_3 (Shah et al., 2023; Sarwar et al., 2024). This process is not included in the chemical mechanism which could contribute to low biases in O_3 during the same time of year. The potential for misattribution is not specific to the methods employed here but is inherent to correlative bias approaches with incomplete information contained in independent variables.

Analysis of the original bias and residual bias emphasize the importance of subpopulation diversity. The correction factors are optimized for the whole population and can degrade performance at any subpopulation (e.g., a site, a day, or a subgroup). For example, in the western US, the PA simulation was originally high-biased for days with high predictions and low-biased for days with high observations (>70 ppb). The overall correction was downward for both populations because they are generally consistent spatially and seasonally. This means that the "corrected" model has more bias on days with high observations in the western US than the "uncorrected." This is not unexpected but highlights that correlative adjustments should be considered as broad conclusions and should only be cautiously applied more narrowly (e.g., specific monitors or days). This is a limitation of the linear formulation as noted by Hosseinpour et al. (2024).

This work has focused only on surface O_3. We are not able to draw a conclusion as to whether the potential lack of stratospheric O_3 is a result of biases in the UTLS PV scaling in the upper layers or from

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

errors in vertical transport from upper layers to the surface. More detailed studies that analyze the entire vertical structure, such as a recent study of CMAQ stratospheric O_3 by Itahashi et al. (2020), are needed to identify the exact causes and solutions for the surface biases identified here. Another potential area for future work is to separate stratospheric O_3 from natural sources in a set of simulations like those conducted for the O_3 Policy Assessment. This might solve the suspected issue of bias in stratospheric O_3 being allocated to long-range international emissions that may be caused by the correlation of stratospheric O_3 and long-range international impacts. Additional future work could take a process-oriented approach rather than the source-oriented approach described here. A process-oriented approach would focus on how different physical and chemical processes (deposition, transport, photochemical activity, etc.) relate to biases in O_3 simulations. A further area for future work is to apply the data fusion bias correction method to an ensemble of USB O_3 estimates from different models. This work has only used the CMAQ model. A test of the method would be to apply it to several different models to determine whether it is able to reduce the uncertainty of USB O_3 estimates while also reducing bias in total O_3.

Acknowledgements

TNS and AGR received funding from the Phillips 66 Company. AGR also received funding from NASA HAQAST. The views expressed in this paper are those of the authors and do not necessarily represent the view or policies of the U.S. Environmental Protection Agency. We thank Benjamin Murphy and Sergey Napelenok for their comments on a draft version of the paper.

Code and data availability

The CMAQ source code is available from GitHub (https://github.com/USEPA/CMAQ) and Zenodo (https://zenodo.org/doi/10.5281/zenodo.1079878). Ozone observational data are available via the AQS website (https://www.epa.gov/aqs).

https://doi.org/10.5194/egusphere-2024-554
Preprint. Discussion started: 25 March 2024
© Author(s) 2024. CC BY 4.0 License.

Author contributions

TNS: conceptualization, investigation, methodology, software, visualization, writing – original draft. CH: data curation, software, writing – review and editing. BHH: data curation, software, writing – review and editing. RM: software, writing – review and editing. KMF: data curation, software, writing – review and editing. AGR: conceptualization, methodology, resources, supervision, writing – review and editing.

Competing interests

The authors declare that they have no competing interests.

References

REFERENCES

Hemispheric Transport of Air Pollution (HTAP). Hemispheric Transport of Air Pollution 2010, Part A: Ozone and Particulate Matter. Task Force on Hemispheric Transport of Air Pollution. Dentener F, Keating T and Akimoto H (eds.). Air Pollution Studies, No. 17 Geneva: United Nations Economic Commission for Europe., 2010.

Appel, K. W., Bash, J. O., Fahey, K. M., Foley, K. M., Gilliam, R. C., Hogrefe, C., Hutzell, W. T., Kang, D., Mathur, R., Murphy, B. N., Napelenok, S. L., Nolte, C. G., Pleim, J. E., Pouliot, G. A., Pye, H. O. T., Ran, L., Roselle, S. J., Sarwar, G., Schwede, D. B., Sidi, F. I., Spero, T. L., and Wong, D. C.: The Community Multiscale Air Quality (CMAQ) model versions 5.3 and 5.3.1: system updates and evaluation, Geosci. Model Dev., 14, 2867-2897, 10.5194/gmd-14-2867-2021, 2021.

Dolwick, P., Akhtar, F., Baker, K. R., Possiel, N., Simon, H., and Tonnesen, G.: Comparison of background ozone estimates over the western United States based on two separate model methodologies, Atmospheric Environment, 109, 282-296, https://doi.org/10.1016/j.atmosenv.2015.01.005, 2015.

Fiore, A., Jacob, D. J., Liu, H., Yantosca, R. M., Fairlie, T. D., and Li, Q.: Variability in surface ozone background over the United States: Implications for air quality policy, 108, 10.1029/2003jd003855, 2003.

Fiore, A. M., Oberman, J. T., Lin, M. Y., Zhang, L., Clifton, O. E., Jacob, D. J., Naik, V., Horowitz, L. W., Pinto, J. P., and Milly, G. P.: Estimating North American background ozone in U.S. surface air with two independent global models: Variability, uncertainties, and recommendations, Atmospheric Environment, 96, 284-300, https://doi.org/10.1016/j.atmosenv.2014.07.045, 2014.

Foley, K., Pouliot, G., Eyth, A., Possiel, N., Aldridge, M., Allen, C., Appel, W., Bash, J., Beardsley, M., Beidler, J., Choi, D., Eder, B., Farkas, C., Gilliam, R., Godfrey, J., Henderson, B., Hogrefe, C., Koplitz, S., Mason, R., Mathur, R., Misenis, C., Pye, H., Reynolds, L., Roark, M., Roberts, S., Schwede, D., Seltzer, K., Sonntag, D., Talgo, K., Toro, C., and Vukovich, J.: EQUATES: EPA's Air QUAlity TimE Series Project, 19th Annual CMAS Conference 2020.

Foley, K. M., Pouliot, G. A., Eyth, A., Aldridge, M. F., Allen, C., Appel, K. W., Bash, J. O., Beardsley, M., Beidler, J., Choi, D., Farkas, C., Gilliam, R. C., Godfrey, J., Henderson, B. H., Hogrefe, C., Koplitz, S. N., Mason, R., Mathur, R., Misenis, C., Possiel, N., Pye, H. O. T., Reynolds, L., Roark, M., Roberts, S., Schwede, D. B., Seltzer, K. M., Sonntag, D., Talgo, K., Toro, C., Vukovich, J., Xing, J., and Adams, E.: 2002–2017 anthropogenic emissions data for air quality modeling over the United States, Data in Brief, 109022, https://doi.org/10.1016/j.dib.2023.109022, 2023.

Guo, J. J., Fiore, A. M., Murray, L. T., Jaffe, D. A., Schnell, J. L., Moore, C. T., and Milly, G. P.: Average versus high surface ozone levels over the continental USA: model bias, background influences, and interannual variability, Atmos. Chem. Phys., 18, 12123-12140, 10.5194/acp-18-12123-2018, 2018.

Hogrefe, C., Liu, P., Pouliot, G., Mathur, R., Roselle, S., Flemming, J., Lin, M., and Park, R. J.: Impacts of different characterizations of large-scale background on simulated regional-scale ozone over the continental United States, Atmos. Chem. Phys., 18, 3839-3864, 10.5194/acp-18-3839-2018, 2018.

© Author(s) 2024. CC BY 4.0 License.

Hosseinpour, F., Kumar, N., Tran, T., and Knipping, E.: Using machine learning to improve the estimate of U.S. background ozone, Atmospheric Environment, 316, 120145, https://doi.org/10.1016/j.atmosenv.2023.120145, 2024.

Huang, M., Bowman, K. W., Carmichael, G. R., Lee, M., Chai, T., Spak, S. N., Henze, D. K., Darmenov, A. S., and da Silva, A. M.: Improved western U.S. background ozone estimates via constraining nonlocal and local source contributions using Aura TES and OMI observations, 120, 3572-3592, 10.1002/2014jd022993, 2015.

Itahashi, S., Mathur, R., Hogrefe, C., and Zhang, Y.: Modeling stratospheric intrusion and trans-Pacific transport on tropospheric ozone using hemispheric CMAQ during April 2010 – Part 1: Model evaluation and air mass characterization for stratosphere–troposphere transport, Atmos. Chem. Phys., 20, 3373-3396, 10.5194/acp-20-3373-2020, 2020.

Jaffe, D. A., Cooper, O. R., Fiore, A. M., Henderson, B. H., Tonnesen, G. S., Russell, A. G., Henze, D. K., Langford, A. O., Lin, M. Y., and Moore, T.: Scientific assessment of background ozone over the US: Implications for air quality management, Elementa-Sci. Anthrop., 6, 30, 10.1525/elementa.309, 2018.

Lin, M., Fiore, A. M., Horowitz, L. W., Langford, A. O., Oltmans, S. J., Tarasick, D., and Rieder, H. E.: Climate variability modulates western US ozone air quality in spring via deep stratospheric intrusions, Nature Communications, 6, 7105, 10.1038/ncomms8105, 2015.

Liu, S. C., Trainer, M., Fehsenfeld, F. C., Parrish, D. D., Williams, E. J., Fahey, D. W., Hübler, G., and Murphy, P. C.: Ozone production in the rural troposphere and the implications for regional and global ozone distributions, Journal of Geophysical Research: Atmospheres, 92, 4191-4207, https://doi.org/10.1029/JD092iD04p04191, 1987.

Mathur, R., Kang, D., Napelenok, S. L., Xing, J., Hogrefe, C., Sarwar, G., Itahashi, S., and Henderson, B. H.: How Have Divergent Global Emission Trends Influenced Long-Range Transported Ozone to North America?, Journal of Geophysical Research: Atmospheres, 127, e2022JD036926, https://doi.org/10.1029/2022JD036926, 2022.

Mathur, R., Xing, J., Gilliam, R., Sarwar, G., Hogrefe, C., Pleim, J., Pouliot, G., Roselle, S., Spero, T. L., Wong, D. C., and Young, J.: Extending the Community Multiscale Air Quality (CMAQ) Modeling System to Hemispheric Scales: Overview of Process Considerations and Initial Applications, Atmos Chem Phys, 17, 12449-12474, 10.5194/acp-17-12449-2017, 2017.

McDonald-Buller, E. C., Allen, D. T., Brown, N., Jacob, D. J., Jaffe, D., Kolb, C. E., Lefohn, A. S., Oltmans, S., Parrish, D. D., Yarwood, G., and Zhang, L.: Establishing Policy Relevant Background (PRB) Ozone Concentrations in the United States, Environmental Science & Technology, 45, 9484-9497, 10.1021/es2022818, 2011.

Sarwar, G., Hogrefe, C., Henderson, B. H., Mathur, R., Gilliam, R., Callaghan, A. B., Lee, J., and Carpenter, L. J.: Impact of particulate nitrate photolysis on air quality over the Northern Hemisphere, Science of The Total Environment, 917, 170406, https://doi.org/10.1016/j.scitotenv.2024.170406, 2024.

Sarwar, G., Gantt, B., Foley, K., Fahey, K., Spero, T. L., Kang, D., Mathur, R., Foroutan, H., Xing, J., Sherwen, T., and Saiz-Lopez, A.: Influence of bromine and iodine chemistry on annual, seasonal, diurnal, and background ozone: CMAQ simulations over the Northern Hemisphere, Atmospheric Environment, 213, 395-404, https://doi.org/10.1016/j.atmosenv.2019.06.020, 2019.

Shah, V., Jacob, D. J., Dang, R., Lamsal, L. N., Strode, S. A., Steenrod, S. D., Boersma, K. F., Eastham, S. D., Fritz, T. M., Thompson, C., Peischl, J., Bourgeois, I., Pollack, I. B., Nault, B. A., Cohen, R. C., Campuzano-Jost, P., Jimenez, J. L., Andersen, S. T., Carpenter, L. J., Sherwen, T., and Evans, M. J.: Nitrogen oxides in the free troposphere: implications for tropospheric oxidants and the interpretation of satellite NO2 measurements, Atmos. Chem. Phys., 23, 1227-1257, 10.5194/acp-23-1227-2023, 2023.

Skipper, T. N., Hu, Y., Odman, M. T., Henderson, B. H., Hogrefe, C., Mathur, R., and Russell, A. G.: Estimating US Background Ozone Using Data Fusion, Environmental Science & Technology, 55, 4504-4512, 10.1021/acs.est.0c08625, 2021.

USEPA: Integrated Science Assessment (ISA) of Ozone and Related Photochemical Oxidants (Final Report, Feb 2013). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-10/076F, 2013.

USEPA: Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards. U.S. Environmental Protection Agency, Washington, DC, EPA-452/R-14/006., 2014.

USEPA: Technical Support Document (TSD) Preparation of Emissions Inventories for the Version 7.1 2016 North American Emissions Modeling Platform, 2019.

USEPA: Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants (Final Report). U.S. Environmental Protection Agency. Washington, DC. EPA/600/R-20/012, 2020a.

USEPA: Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards. U.S. Environmental Protection Agency, Washington, DC, EPA-452/R-20-001, 2020b.

Xing, J., Mathur, R., Pleim, J., Hogrefe, C., Wang, J., Gan, C. M., Sarwar, G., Wong, D. C., and McKeen, S.: Representing the effects of stratosphere–troposphere exchange on 3-D O3 distributions in chemistry transport models using a potential vorticity-based parameterization, Atmos. Chem. Phys., 16, 10865-10877, 10.5194/acp-16-10865-2016, 2016.