***Average versus high surface ozone levels over the continental U.S.A.: Model bias, background influences, and interannual variability*, Jean Guo et al., ACP, (2018)**

Authors' response is written in bold type; Reviewer comment in normal type. Figures, including new ones added to the paper that were not in direct response to reviewer's comments are at the end.

Author response to Reviewer #1

This manuscript presents an attempt to derive information about mean maximum daily 8-hour average (MDA8) O_3 in the United States, based on ambient measurements and using the global model GEOS-CHEM. Sensitivity simulations examine different sources that affect the 10 highest O_3 events and that affect the 10 days with highest model bias against observations for 2004 to 2012 for each 10 EPA regions.

General comments: The analysis is a valuable contribution to the current understanding of ground level O_3 and air quality standard settings. The topic itself is highly relevant and thus will be of interest to the readers of ACP. Discussion of the results and their implications is also scientifically sound and the paper includes comprehensive analyses. However, I feel that the paper tried to cover lots of information, which makes it a bit hard for the reader to follow key conclusions from this study. Thus, I recommend that the paper should be published after addressing the following comments.

General

Some comment about day of week effects and model biases in temperature as they relate to the questions raised in the paper seem warranted. There should be comment dramatic changes in the temperature dependence of ozone over this period coincident with the NOx changes. Those changes should have a day of week variation that might appear in the top 10 days.

We have addressed the biases in temperature by adding a comparison to the Global Historical Climatology Network Global Historical Climatology Network (GHCN) and the Climate Anomaly Monitoring System (CAMS). See Supplemental Table 4 and the associated discussion in the text (lines 363-364): *"The model monthly mean temperatures in the model (from the MERRA reanalysis) closely match the observed GHCN+CAMS dataset (Supplemental Table 1)."*

The request to investigate day of week effects substantially widens the scope of the paper. The general comments suggest that the manuscript is already covering too much information. We feel that tackling day of week effects and its changes over time is a study unto itself and thus outside the scope of this particular paper.

Specific comments: The authors use terms "Baseline O_3" and "U.S. background O_3". U.S. background O_3 is defined as "the O_3 levels that would exist in the absence of U.S. anthropogenic emissions of precursors" and Baseline O_3 is defined as "tropospheric O_3 concentrations that have a negligible influence from local anthropogenic emissions". They sound the same, don't they? If yes, please be consistent in the text.

These definitions are not the same. Clarification has been added in lines 105-107. *"Baseline O_3 is a measurable quantity and differs from background O_3 in that it contains some influence from U.S. anthropogenic emissions that were not recently emitted but contributed to the global background."* We follow here the definitions of Jaffe et al., 2018 which builds on the 2009 National Academies report "Global Sources of Local Pollution", and the HTAP 2010 report (available at www.htap.org).

Page 4, lines 106-109, Please clarify if the authors apply Schnell et al. (2014)'s interpolation procedure or they use their dataset. Schnell et al. (2014) use surface MDA8 O_3 measurements from air quality networks for 2000–2009, while this paper analyzes the data from 2004-2012.

Jordan Schnell is a co-author and provided the dataset that we used here. The interpolation procedure for his dataset is described in his 2014 paper; he provided us with the data for the years since 2009. We have edited the citation to reflect a newer paper in which this extended dataset has been used. See lines 118-120: *"we use an available 1° x 1° grid of surface MDA8 O_3 measurements that were interpolated from the AQS, CASTNet, and Canadian NAPS networks (Schnell and Prather, 2017)."*

A valuable addition would be a statement about the chemistry scheme applied in the version GEOS-Chem at the 2.3 section (GEOS-Chem model simulation). The authors mention issues of isoprene chemistry in last paragraph of Conclusions but a brief description or reference to the specific version of the chemistry should be presented before the last paragraph of the paper.

On lines 158-160 we now state: *"We use the standard v9_02 chemical mechanism which includes recycling of isoprene nitrates (Mao et al., 2013) in contrast to the mechanisms used in earlier versions of GEOS-Chem (e.g., Zhang et al., 2014 as discussed in Fiore et al., 2014)."*

The last paragraph of page 6 needs elaboration where the authors state the sensitivity simulations. The notations for all model simulations should be mentioned and the description of Table 1 should be modified so that the Table is read from top to bottom. **We have completely rewritten this section with the intent of improving clarity. See paragraph starting from line 183 *("We first perform a base simulation…")*.**

Figure 3: Observed O_3 concentrations should be represented in a different color to be more visible (maybe black instead of grey) and I would also suggest to plot the curves as an average for 2004-2012 period with associated error bars. **Thanks for this suggestion. We have edited the figure to show the curves as an average for 2004-2012 period with associated error bars.**

Minor comments: The tables start from Table 2 at the manuscript and Table 1 is referenced at Page 8 for the first time. Please fix ordering of table numbers as they appear in the text. **Fixed**

Page 7, line 195: "a maximum in and" should read "a maximum in summertime and" **Thank you. Section has been edited and this sentence was removed.**

Author response to Reviewer #2

General:

The paper is very well-written and concerns a topic of considerable interest to air quality planners. However, there are some concerns about the suitability of this particular model configuration to address some of the stated objectives of the paper (lines 76-79), as discussed below. In general, the paper would be improved if there was greater clarity about the potential connections between the findings and possible configuration concerns. The value of the paper would be enhanced if the conclusions section was bolstered with a "next steps" or "considerations" sentence or two that described how such a global model-based sensitivity study could be improved in the future.

We have attempted to strengthen the paper throughout as suggested by the reviewer. In particular, we added a sentence in the introduction, (lines 77-80) to highlight the key benefit and drawback of using a coarse resolution model: *"Though coarse resolution global models such as GEOS-Chem will mix emissions into the same grid cell that may remain separate in the real atmosphere, a global model is necessary to quantify background O_3 transported intercontinentally, including that produced via oxidation of methane."*

We added a sentence in the conclusion (lines 476-481) to emphasize a need to confirm our findings with finer scale models: *"Future work with high-resolution models (e.g., at the regional scale, ideally with boundary conditions that include source attributions from a global model) is needed, along with observational evidence, to quantify the extent to which biogenic VOC and NOx contribute to the highest observed O_3 levels in the warm season. The importance of temperature sensitive sources like biogenic VOC and NOx emissions to background O_3 imply that in a warmer climate, these background influences on O_3 will play an even more important role in driving up O_3 levels."*

In particular, there is concern about the use of a coarse resolution model (2 x 2.5 deg) to investigate contributions of U.S. anthropogenic emissions (O_3_USA) given that those contributions originate at scales much smaller than the resolution of the model (i.e., point source emissions, urban area emissions). The paper acknowledges the limitations associated with the coarse modeled resolution in several places (lines 212, 242, 399). The paper may want to revisit these caveats in the conclusion and perhaps provide some thoughts on what alternate global model configurations would be better suited for an analysis of source contributions.

Agreed. Please see above.

Kudos to the authors for providing sufficient detail regarding the performance evaluation to allow readers to interpret the contribution findings in light of the model bias/error. However, the ozone overestimations (3-14 ppb in JJA MDA8 top 10 days by region, even worse for JJA all-day averages) suggest caution should be exercised in overinterpreting the contributions. Based on Figure 5 and the associated analyses, it appears that the model vastly overestimates ozone on hot days in the late summer, especially in the eastern U.S. (even without consideration of potential additional emissions due to increased power demand on those days). Section 3.3 briefly

summarizes potential causes for this overestimation based on similar studies, but it would be valuable if the paper provided more application-specific hypotheses for the underlying cause.

We have added more discussion of the potential for biases in the meteorology (see responses to reviewer #3), as well as in anthropogenic NO_x emissions to contribute to the summertime overestimate in the model compared to observations. As our set of sensitivity simulations identifies a potential role for biogenic VOC and soil NO_x in contributing to the bias, we have added to the text some discussion calling out the need for better constraints on these biogenic emissions, though we do note that the model nevertheless shows some skill at capturing the observed year-to-year variability, which includes a correlation with O_3 produced from natural sources (BVOC and soil NO_x), which, like total O_3, correlate with temperature (Figure 6).

We now state, in lines 441-449 ***"Our finding that BVOC emissions contribute to the summertime surface O_3 biases could reflect poor representation of the emissions (and subsequent oxidation chemistry). Earlier work has noted that MEGAN BVOC emissions are too high over California (Bash et al., 2016), Southeast Texas (Kota et al., 2015), the Ozarks in southern Missouri (Carlton and Baker, 2011), and across much of the U.S.A. (Wang et al., 2017). One recent model study uniformly reduced MEGAN isoprene emissions by 20% (Li et al., ACP 2018), but we did not apply any such scaling here. In regions that are highly NOx-sensitive, additional isoprene should not strongly influence O_3, as found over southeast Texas (Kota et al., 2015). While not eliminated entirely, the summertime model bias does lessen in the simulation with BVOC emissions set to zero, suggesting that the O_3 bias is indeed exacerbated if BVOC emissions are overestimated in the model."***

FYI, along w/ the possible causes from the Travis research, others have raised concerns about MEGAN biogenic VOC estimates (e.g., Bash et al., 2016; Carlton and Baker, 2011; Kota et al., 2015; Wang et al., 2017). **Thank you for pointing these out. We added these references (see previous response).**

One of the more noteworthy findings concerns the modeled trends over the 10-year period (e.g., lines 386-389) where the analysis appears to confirm previous findings that improving trends in U.S. air quality from emissions controls have been tempered by increases in background contributions (and increases in temperature). However, one interesting finding here that could use additional explanation is the regional breakout of this "USB vs. USA" tradeoff. Table 5 suggests that the largest increases in high JJA-day O_3_USB concentrations between 2004-2006 and 2010-2012 have occurred in the New England and Mid-Atlantic regions, not the western regions where USB concerns are typically greatest. More explanation of the regional differences in modeled USB trends would be beneficial (e.g., is this just an artifact of the meteorology of the two 3-year periods in these regions).

We agree. The "trends" over such a short period are strongly influenced by fluctuations in temperature. While it may indeed be an 'artifact' of looking at such a short period, it nevertheless suggests that regionally produced background O_3 from temperature-sensitive emissions (BVOC and NO_x) may grow in importance in the coming decades in light of a

warming climate. We have attempted to make this clearer by adding a column to Table 4 that shows the change in temperature between these two 3-year periods in each region. We have edited the accompanying discussion to the main text:

Starting from line 364: ***"Table 4 shows that regions with O_3_USB increases generally experienced rising temperatures over this period, as the 2010-2012 period includes two of the warmest years on record. Figure 6 shows that O_3_NAT tracks with…"***

In response to a comment from Reviewer 1, we have also added Supplementary Table 4 that evaluates monthly mean model temperatures with the Global Historical Climatology Network.

Given model performance findings, would the authors see value in revising the "2-step" contribution analysis (assessing contributions on high-bias days, then assessing contributions on high/all observed days) to a "3-step" contribution where as an intermediate step you also investigated contributions on top-10 modeled days? This could be valuable presuming that the subset of days would differ from top 10 highest bias days.

Though we did conduct some exploratory analysis using this 3-step method early on, we did not end up pursuing this method in the paper because the highest model days are less relevant to the "real world" and if this method were used throughout our paper, the number of figures would have doubled. As the paper is already lengthy, we choose to focus on the days in the observations when the O_3 NAAQS is most likely to be exceeded.

We have, however added text that clarifies the extent to which there is overlap between the highest 10 days in the model and the 10 days with the highest biases: ***"There is at most a 2-6 day overlap between the top 10 O_3_Base days and the top 10 most biased days in 2004-2012 across all regions, but during most years, the overlap is around 0-2 days. We restrict our analysis to examining the top 10 observed O_3 days as these days are most relevant from a policy perspective."*** **(lines 137-139).**

Rather than lumping the Mount Bachelor observations (and subsequent pairs) with surface sites in Region 10, it would be interesting to see how model contributions varied as functions of model performance and observation concentration as a standalone site.

Thanks for this suggestion. We now include a more detailed analysis of Mount Bachelor as a separate standalone section (Section 3.2) that includes new figures (Figure 3 and Supplemental Figure 4).

<u>Specific:</u>

Line 86: "download" should be "downloaded". **FIXED**

Lines 124-127: Would be easier to read, if a new sentence was started w/ "On the days with . . .". **FIXED**

Line 146: "Avian" should be "Aviation". **FIXED**

Line 195: Is the word "summer" missing from this sentence . . . "The model, however, has a maximum in [summer] and underestimates springtime baseline O_3"? **Thank you. Section has been edited and this sentence was removed**

Line 205: Are Travis et al. (2016) conclusions regarding 2011 NEI relevant to a model configuration based on 2005 NEI w/ annual scalars?

We add a comparison with Travis et al. (2016) in lines 226-230: ***"Travis et al. (2016) find that the 3.5 Tg N y-1 NEI 2011 estimate for U.S. fuel NOx emissions is too high and contributes to excessive surface O_3. Our simulations include even higher U.S. fuel NOx emissions of 4.4 Tg N y-1 during 2010-2012 (Supplemental Table 3), implying that some portion of the model O_3 bias reflects excessively high anthropogenic NOx emissions (Travis et al., 2016)."***

Line 248: Clarify that these monthly averages are MDA8 O_3 (not hourly)? **FIXED**

Line 367: Move mention of lack of daily variation in emissions to early section? **Done. Now at lines 179-183)**

Line 396: Same as above, maybe mention this earlier in modeling methodology section? **See lines 179-183)**

<u>Author response to Reviewer #3</u>

This paper presents a comprehensive modeling analysis of surface ozone and the various factors that contribute to its variability over the United States. By conducting multiple sensitivity simulation removing various sources for the 2004-2012 period, the authors estimate the influence of different background sources and of U.S. anthropogenic sources on mean surface O_3 and high O_3 events as a function of region, season, and year.

Two aspects of the paper that I'd like to see more discussion on are listed below:

1) The paper is very detailed with many figures and tables and is one more study on top of a rich set of published work, including by some of the co-authors. The authors often cite previous work, saying it is consistent with their results, but it would be useful to highlight what are the new key contributions from their specific analysis. What new information did the detailed modeling analysis bring to this prolific field?

We added text with the intent of providing stronger motivation to the introduction in which we highlight the use of sensitivity simulations to help us identify which sources contribute most to the summertime bias and to the highest O_3 days (lines 70-87). To our knowledge, the finding that increasing O_3 production from temperature-sensitive biogenic emissions might be offsetting some of the gains achieved by reducing anthropogenic ozone precursor emissions is new, and potentially of growing importance as record-setting warm years have been increasing. We believe that our finding that the summertime bias is associated with regionally produced ozone – including both U.S. anthropogenic and components of U.S.

background – rather than transported background (either internationally or intercontinentally) is also new. We have also rewritten the conclusions to emphasize these points.

2) There isn't much discussion on the causes of the large summer bias over the Eastern US and how this bias affects the interpretation of the results.

To our knowledge, prior studies have not used such a broad set of sensitivity simulations to interpret which sources are contributing most in places and times when the model is most biased against observations. Section 3.3 in the submitted paper is entirely devoted to addressing this point. We thus assume that the reviewer is instead driving at the deeper question of the specific causes of the bias, beyond what we can identify cleanly with the sensitivity simulations. We have added additional discussion in response to reviewer 2 that attempts to address both the causes and how it affects the interpretation of the results.

Specifically, we added

1) **A sentence in the introduction, (lines 77-80) to highlight the key benefit and drawback of using a coarse resolution model: *"Though coarse resolution global models such as GEOS-Chem will mix emissions into the same grid cell that may remain separate in the real atmosphere, a global model is necessary to quantify background O_3 transported intercontinentally, including that produced via oxidation of methane."***
2) **We also added a sentence in the conclusion (lines 476-481) to emphasize a need to confirm our findings with finer scale models: "*Future work with high-resolution models (e.g., at the regional scale, ideally with boundary conditions that include source attributions from a global model) is needed, along with observational evidence, to quantify the extent to which biogenic VOC and NOx contribute to the highest observed O_3 levels in the warm season. The importance of temperature sensitive sources like biogenic VOC and NOx emissions to background O_3 imply that in a warmer climate, these background influences on O_3 will play an even more important role in driving up O_3 levels.*"**
3) **We now state, in lines 441-449 *"Our finding that BVOC emissions contribute to the summertime surface O_3 biases could reflect poor representation of the emissions (and subsequent oxidation chemistry). Earlier work has noted that MEGAN BVOC emissions are too high over California (Bash et al., 2016), Southeast Texas (Kota et al., 2015), the Ozarks in southern Missouri (Carlton and Baker, 2011), and across much of the U.S.A. (Wang et al., 2017). One recent model study uniformly reduced MEGAN isoprene emissions by 20% (Li et al., ACP 2018), but we did not apply any such scaling here. In regions that are highly NOx-sensitive, additional isoprene should not strongly influence O_3, as found over southeast Texas (Kota et al., 2015). While not eliminated entirely, the summertime model bias does lessen in the simulation with BVOC emissions set to zero, suggesting that the O_3 bias is indeed exacerbated if BVOC emissions are overestimated in the model."***

Discussing this in more detail would strengthen the paper. The authors have one sentence addressing this by referring to the work of Travis et al. (2016) using a more recent version of the GEOS-Chem model. They mention potential errors in anthropogenic NOx emission in the NEI inventory, but Travis et al. use the NEI 2011 inventory while the authors use the NEI 2005 inventory. How different are they? If the NEI NOx inventory is indeed too high, how would that affect the calculation of O_3_USA?

We now directly compare our NOx emissions to those used in Travis et al., 2016 and include a supplementary table (see above – Supplemental Table 3) providing the NOx emissions applied in each year within the U.S.A. and globally.

Lines 226-230: "*Travis et al. (2016) find that the 3.5 Tg N y-1 NEI 2011 estimate for U.S. fuel NOx emissions is too high and contributes to excessive surface O_3. Our simulations include even higher U.S. fuel NOx emissions of 4.4 Tg N y-1 during 2010-2012 (Supplemental Table 3), implying that some portion of the model O_3 bias reflects excessively high anthropogenic NOx emissions (Travis et al., 2016).*"

They mention meteorological factors associated with boundary layer mixing and cloud cover which would affect the vertical distribution of O_3, but Travis et al. used different meteorological fields (GEOS-FP) compared to the MERRA fields used by the authors. It is unclear whether these potential explanations apply in this case. If MERRA meteorology is indeed biased, then that would certainly affect the validity of the relative influence of various sources on the "most-biased" days analysis and on the average MDA8 O_3 levels. A discussion of this would be valuable.

Thanks for this suggestion. We have attempted to address this point by including more discussion of published evaluations of MERRA meteorology:

1) **Lines 152-157: "*MERRA meteorology captures summer mean surface temperatures to within 1-2 K across U.S. regions and precipitation to within 0.5 mm d-1 except for over the Northern Great Plains where a positive bias exceeds 1 mm d-1, but the variance in summer mean precipitation is lower than observed in some regions (Bosilovich, 2013). While interannual variability in cloudiness observed at weather stations is largely captured by MERRA, the reanalysis generally underestimates cloud cover and thus overestimates observed downward surface shortwave fluxes (Free et al., 2016).*"**
2) **Lines 230-232: "*The low bias in cloud cover in the MERRA meteorology and associated overestimate in downward shortwave surface radiation (Free et al., 2016) may also contribute to excessive O_3 production in the model.*"**
3) **We also added our own evaluation of surface temperature over the U.S.A. in the MERRA fields (Supplemental Table 4).**

Minor comments:

Line 154. "Anthropogenic emissions. . . are scaled each year on the basis of economic data". It would be useful to have a bit more discussion on how anthropogenic emissions are scaled over

the continental U.S. which uses 2005 as the baseline. By how much do NOx emissions change over the time period of the simulation 2004-2012.

Supplemental Table 3 was added to provide the NOx emissions within each year, both globally and within the U.S.A. (Lines 178-183)

Are these scaling factors taken from the NEI trends report (https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data) itself or was independent estimate done?

The van Donkelaar et al., 2008 describes the standard GEOS-Chem emissions scaling reference. The scale factors use government statistics where available.

Edited this sentence (lines 169-170) to include "provided by individual countries, where available"

Line 195. "a maximum in and underestimate springtime. . ." is "summer" missing after maximum? **Yes. Thank you**

Line 196. While the authors talk about potential causes for the springtime underestimate (stratospheric intrusions), they do not talk about the summertime overestimate, which is quite large.

See response to general comments above and our additions above regarding anthropogenic NO_x emissions and citations of prior work evaluating MERRA meteorology (temperature, precipitation and cloud cover).

Figures and tables edited/added in response to reviewer's comments:

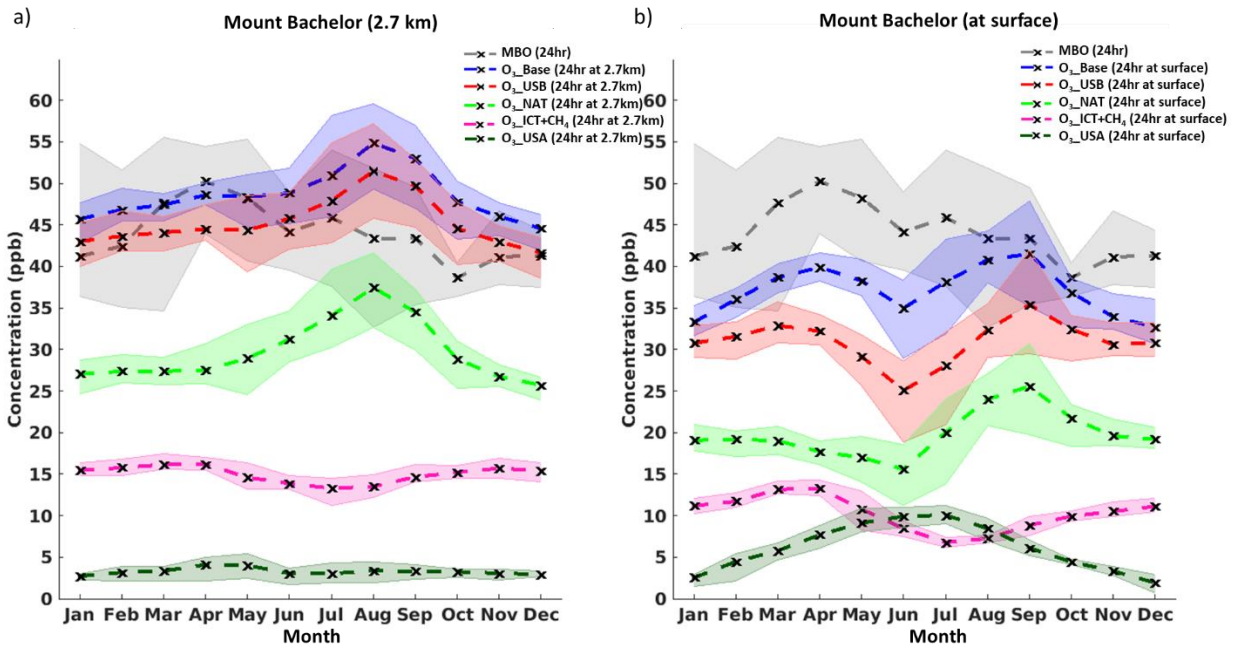
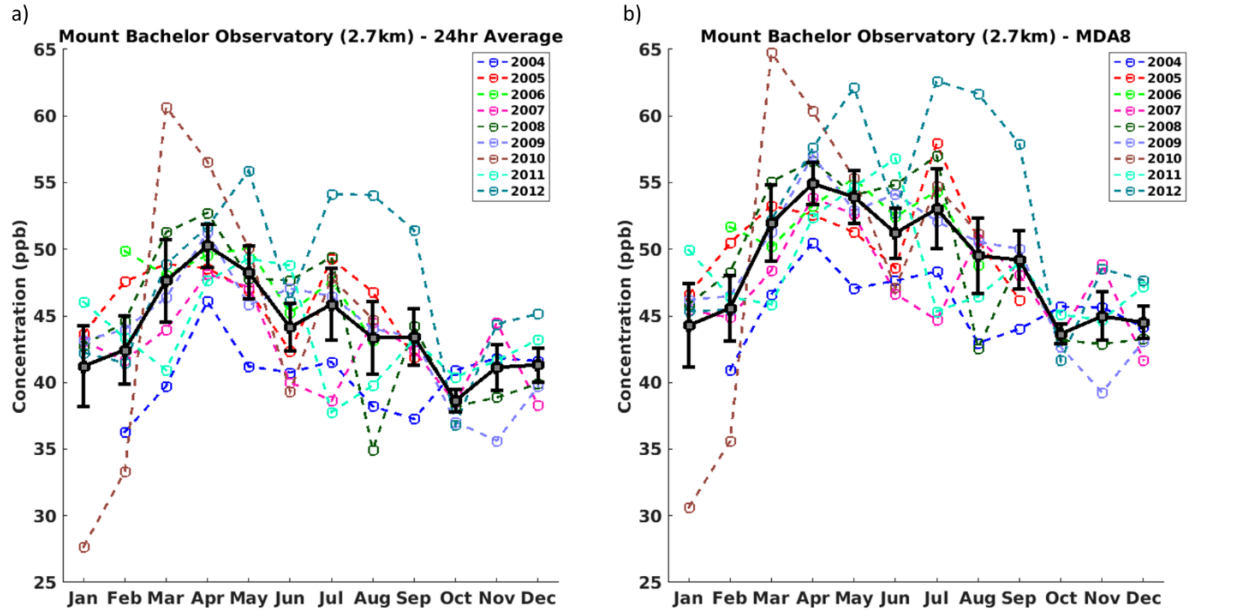


Figure 1: Monthly 2004-2012 average 24-hour O_3 concentrations at Mount Bachelor Observatory. Observations (grey) are the same in both panels. Simulations from the GEOS-Chem model are sampled in the grid cell containing Mount Bachelor at (a) 2.7 km (the height of the Mount Bachelor Observatory) and at (b) the surface: O_3_Base (blue), O_3_USB (red), O_3_NAT (light green), O_3_ICT+CH_4 (pink), and O_3_USA (dark green). The shaded range spans the highest and lowest years.



Supplemental Figure 1: Monthly average of observed (a) daily 24-hour and (b) MDA8 O_3 concentrations averaged across 2004-2012 at Mount Bachelor Observatory. Black line shows the average of each month from 2004-2012. Error bars show the standard deviation in the interannual variability in each month. Dashed lines show the concentrations for each individual year.

Table 1: Change in MDA8 O_3 concentrations from 2004-2006 to 2010-2012 on O_3_top10obs_JJA days in the observations, O_3_Base, O_3_USB, O_3_USA, and temperature.

	Obs	O_3_Base	O_3_USB	O_3_USA	Temperature (C)
New England	-6	-4	6	-10	2
NY+NJ	-2	-4	3	-7	1
Mid-Atlantic	0	-3	4	-7	1
Southeast	-4	-5	2	-7	1
Midwest	-2	-4	2	-6	0
South Central	-6	-2	5	-7	1
Plains	-1	-2	4	-5	1
Mountains + Plains	-4	-1	1	-2	-1
Pacific SW	-3	-4	0	-4	-1
Pacific NW	-7	-5	-4	-1	-1
Average	***-3***	***-3***	***2***	***-6***	***0***

Supplemental Table 1: Monthly average temperature across all days in each season (average of 2004-2012) in (1) GEOS-Chem, in (2) the Global Historical Climatology Network (GHCN) and the Climate Anomaly Monitoring System (CAMS) (in degrees C), and (3) the difference between these values.

	Model Temperature (C)			GHCN+CAMS Temperature (C)			Model Temp. Bias		
Region	***MAM***	***JJA***	***SON***	***MAM***	***JJA***	***SON***	***MAM***	***JJA***	***SON***
New England	7	19	11	8	20	11	0	-1	1
NY+NJ	9	21	12	9	21	12	0	0	1
Mid-Atlantic	12	23	14	12	23	14	0	0	0
Southeast	17	26	18	17	26	18	0	0	0
Midwest	10	22	12	10	22	12	0	0	0
South Central	18	27	19	19	28	20	-1	-1	-1
Plains	13	25	13	13	25	13	0	0	0
Mountains + Plains	7	20	9	7	19	8	0	1	1
Pacific SW	14	23	17	14	22	17	-1	0	0
Pacific NW	8	17	10	8	17	9	0	0	1

Supplemental Table 2: Global and US emissions totals for 2004-2012.

	Emissions	2004	2005	2006	2007	2008	2009	2010	2011	2012
Global	*Anthropogenic NO with biofuels (Tg N)*	30.3	30.2	30.1	29.9	29.5	29.0	28.8	28.8	28.8
	Biomass burning (Tg N)	4.5	4.7	4.6	4.6	3.9	3.5	5.0	3.7	3.7
	Soil (Tg N)	9.0	9.1	8.8	8.5	8.4	8.6	8.4	8.6	9.2
	Lightning (Tg N)	5.5	6.1	6.2	6.4	6.9	7.3	7.2	7.1	7.2
	Isoprene (Tg C)	493.0	499.3	471.5	453.6	435.3	455.4	466.0	453.3	467.3
US	*Anthropogenic NO with biofuels (Tg N)*	6.32	6.04	5.75	5.44	5.13	4.63	4.36	4.36	4.36
	Biomass burning (Tg N)	0.02	0.06	0.06	0.07	0.04	0.04	0.05	0.12	0.12
	Soil (Tg N)	0.78	0.86	1.02	0.92	0.82	0.79	0.77	0.95	1.10
	Lightning (Tg N)	0.86	0.86	0.77	0.75	1.10	1.13	1.13	1.28	1.31
	Isoprene (Tg C)	18.1	21.5	21.9	22.0	19.3	18.3	20.2	22.0	22.4

Additional figures/tables added to paper:

Supplemental Table 3: Correlation between (1) O_3_Base and O_3_USB and (2) O_3_Base and O_3_USA on the average of O_3_top10obs_JJA days from 2004-2012 in each region.

	Correlation	
Region	**O_3 Base and O_3 USB**	**O_3 Base and O_3 USA**
New England	0.28	0.64
NY+NJ	0.50	0.58
Mid-Atlantic	0.54	0.70
Southeast	0.66	0.59
Midwest	0.75	0.76
South Central	0.71	0.72
Plains	0.80	0.75
Mountains + Plains	0.95	0.64
Pacific SW	0.72	0.28
Pacific NW	0.98	0.05

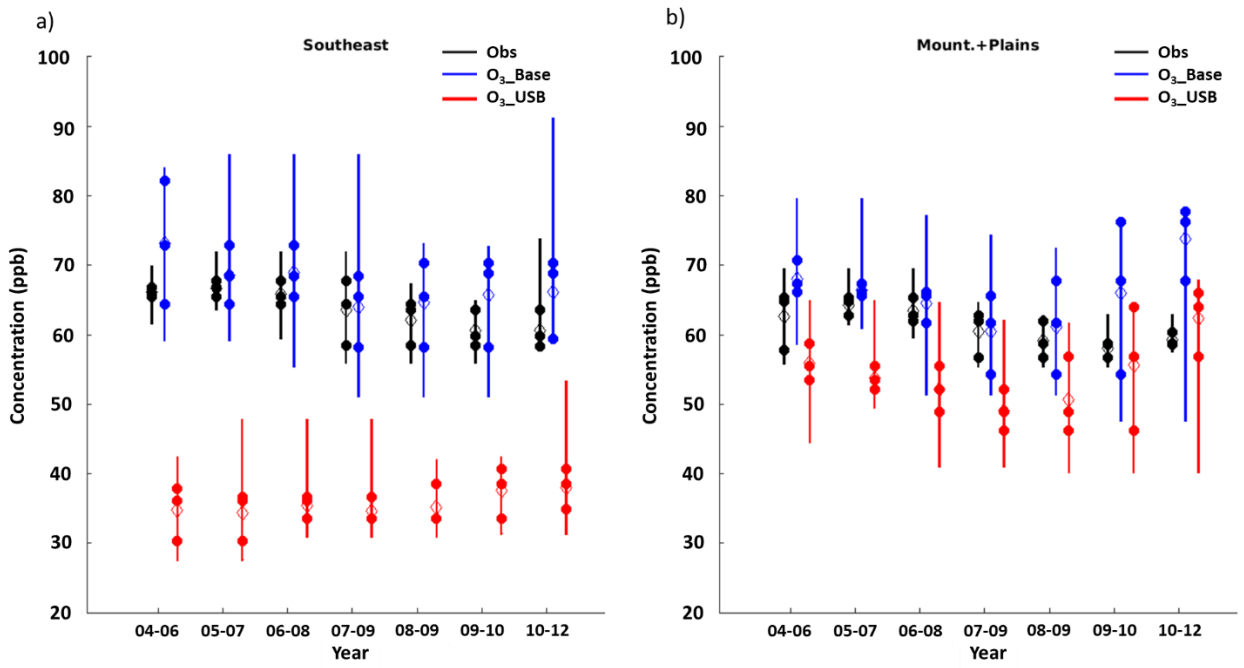
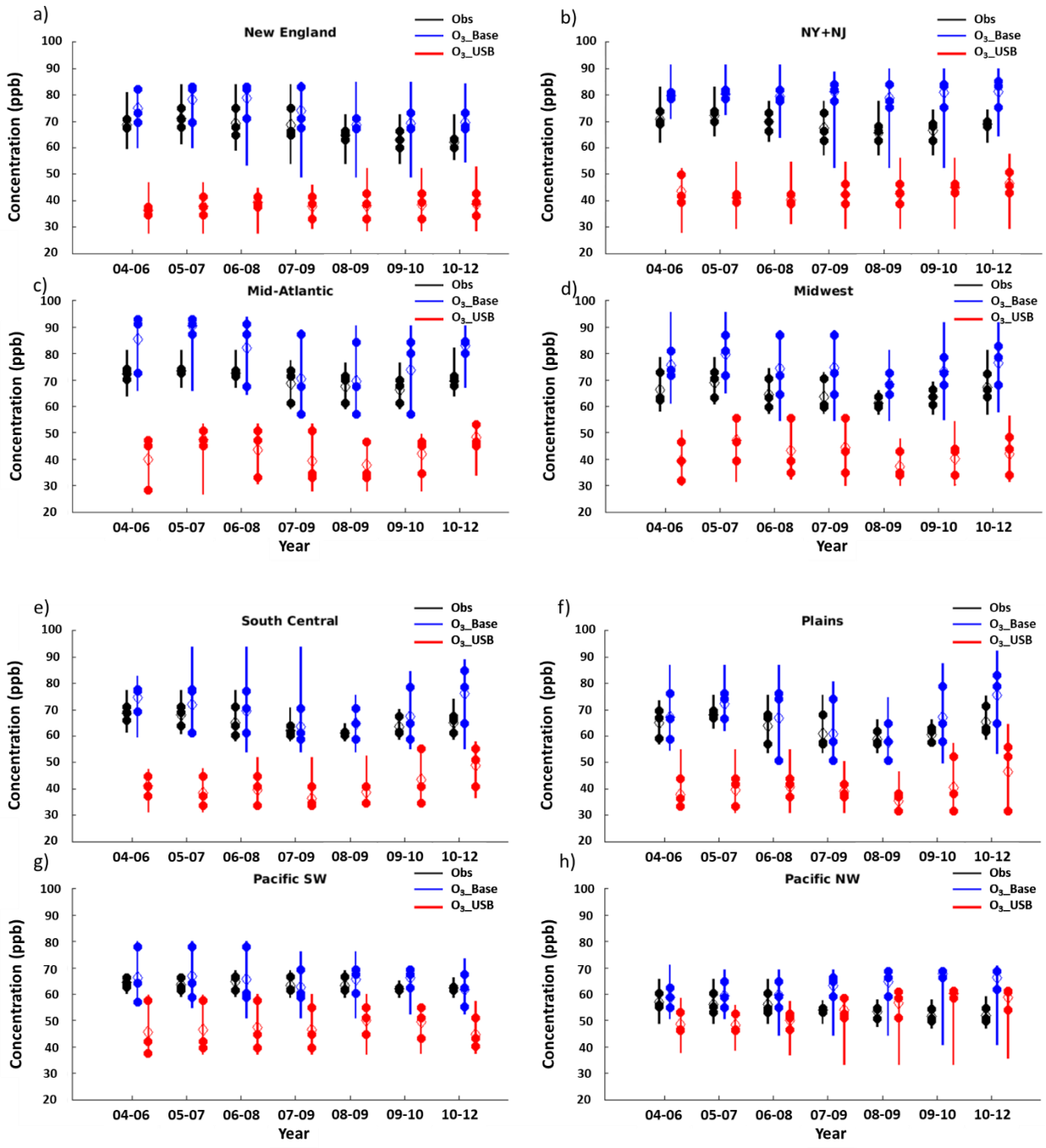


Figure 2: The three 4th highest days in each year (solid dots) that went into the calculation of the three-year average of the 4th highest MDA8 O_3 day (hollow diamond). Error bars show the range between the highest and lowest O_3_top10obs days across each 3-year span (i.e, across 30 total points) occurring between March and October in the (a) Southeast and (b) Mountains and Plains regions in the observations (black), and the O_3_Base (blue) and O_3_USB (red) simulations sampled on the same days as the top 10 observed values.



Supplemental Figure 2: Summary information for each region showing the three 4th highest days in each year (solid dots) that went into the calculation of the three-year average of the 4th highest MDA8 O_3 day (hollow diamond). Error bars show the range between the highest and lowest O_3_top10obs days across each 3-year span (i.e, across 30 total points) occurring between March and October in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW. Observations are shown in black, O_3_Base is in blue, and O_3_USB is in red.

Table 2: Summary information for each region. The first row next to each region reports the range across 2004-2012 of the 4th highest values from each of the 9 individual years for the observations, O_3_Base, and O_3_USB. The second row reports the range across 2004-2012 of each of the 3-year averages of the 4th highest values (7 values) in each region for the observations, O_3_Base, and O_3_USB.

Region	*Range*	Obs	O_3_Base	O_3_USB
New England	4th highest day	15	16	10
	3-year average 4th highest day	9	10	3
	Difference	**-6**	**-6**	**-7**
NY+NJ	4th highest day	11	10	12
	3-year average 4th highest day	6	2	6
	Difference	**-5**	**-8**	**-6**
Mid-Atlantic	4th highest day	13	36	25
	3-year average 4th highest day	7	21	10
	Difference	**-6**	**-15**	**-15**
Southeast	4th highest day	9	24	10
	3-year average 4th highest day	6	9	4
	Difference	**-3**	**-15**	**-7**
Midwest	4th highest day	13	22	24
	3-year average 4th highest day	8	11	10
	Difference	**-6**	**-11**	**-14**
South Central	4th highest day	11	26	22
	3-year average 4th highest day	8	13	13
	Difference	**-3**	**-13**	**-9**
Plains	4th highest day	14	32	24
	3-year average 4th highest day	9	18	11
	Difference	**-5**	**-15**	**-13**
Mountains + Plains	4th highest day	9	23	20
	3-year average 4th highest day	6	13	13
	Difference	**-2**	**-10**	**-7**
Pacific SW	4th highest day	5	23	20
	3-year average 4th highest day	3	5	5
	Difference	**-2**	**-18**	**-15**
Pacific NW	4th highest day	11	14	15
	3-year average 4th highest day	5	9	12
	Difference	**-5**	**-5**	**-3**

Average versus high surface ozone levels over the continental U.S.A.: Model bias, background influences, and interannual variability

Jean J. Guo[1], Arlene M. Fiore[1], Lee T. Murray[2,3,4], Daniel A. Jaffe[5], Jordan L. Schnell[6,7], Tom Moore[8], George Milly[2]

[1]Department of Earth and Environmental Sciences and Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, U.S.A.
[2]Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, U.S.A.
[3]NASA Goddard Institute for Space Studies, New York, NY USA
[4]Now at: Department of Earth and Environmental Sciences, University of Rochester, Rochester, NY, U.S.A.
[5]University of Washington, School of STEM, Bothell, WA and Department of Atmospheric Science, Seattle, WA, U.S.A.
[6]NOAA Geophysical Fluid Dynamics Laboratory, Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, U.S.A.
[7]Now at: Department of Earth and Planetary Sciences, Northwestern University, Chicago, IL, U.S.A.
[8]WESTAR and WRAP, Colorado State University, Fort Collins, CO, U.S.A.

Correspondence to: Jean J. Guo (jean.j.guo@columbia.edu)

Abstract. U.S. background ozone (O_3) includes O_3 produced from anthropogenic O_3 precursors emitted outside of the U.S.A., from global methane, and from any natural sources. Using a suite of sensitivity simulations in the GEOS-Chem global chemistry-transport model, we estimate the influence from individual background versus U.S. anthropogenic sources on total surface O_3 over ten continental U.S. regions from 2004-2012. Evaluation with observations reveals model biases of +0-19 ppb in seasonal mean maximum daily 8-hour average (MDA8) O_3, highest in summer over the eastern U.S.A. Simulated high-O_3 events cluster too late in the season. We link these model biases to excessive regional O_3 production (e.g., U.S. anthropogenic, biogenic volatile organic compounds (BVOC), and soil NO_x, emissions), or coincident missing sinks. On the ten highest observed O_3 days during summer (O_3_top10obs_JJA), U.S. anthropogenic emissions enhance O_3 by 5-11 ppb and by less than 2 ppb in the eastern versus western U.S.A. The O_3 enhancement from BVOC emissions during summer is 1-7 ppb higher on O_3_top10obs_JJA days than on average days, while intercontinental pollution is up to 2 ppb higher on average ~~vs~~versus on O_3_top10obs_JJA days. ~~In the model, regional sources of O_3 precursor emissions drive interannual variability in the highest observed O_3 levels.~~ During the summers of 2004-2012, monthly regional mean U.S. background O_3 MDA8 levels vary by ~~10-20~~up to 15 ppb~~. Simulated~~ from year to year. Observed and simulated summertime total surface O_3 levels on O_3_top10obs_JJA days decline by 3 ppb (averaged over all regions) from 2004-2006 to 2010-2012~~ in both the observations and the model~~, reflecting rising U.S. background (+2 ppb) and declining U.S. anthropogenic O_3 emissions (-6 ppb)~~.~~ in the

model. The model attributes interannual variability in U.S. background O_3 on O_3_top10obs days to natural sources, not international pollution transport. We find that a three-year averaging period is not long enough to eliminate interannual variability in background O_3. on the highest observed O_3 days.

1 Introduction

In the United States, ozone (O_3) is regulated as a criteria pollutant under the National Ambient Air Quality Standard (NAAQS). The current NAAQS for ground-level O_3, set in October 2015, states that the 4th-highest daily maximum 8-hour average (MDA8) O_3, averaged across three consecutive years, cannot be 71 ppb or higher (U.S. Environmental Protection Agency, 2015). The three-year average is nominally intended to smooth out fluctuations in O_3 levels resulting from natural variability in meteorology within the timing constraints of the federal Clean Air Act for air quality planning. As even one ppb of excess O_3 may be enough to push a county out of NAAQS attainment, it is relevant to understand which sources influence the severity and timing of the highest O_3 events. AsSince measured O_3 does not retain a signature of the source from which it was produced, estimates of background O_3 rely on models, ideally evaluated closely with observational valuesobservations, to build confidence in the model capability for source attribution. Here we apply a global chemistry-transport model alongside O_3 observations to examine the highest 10 observed O_3 events, as well as average conditions, to determine which sources are influencing average versus high-O_3 events, and the extent to which they vary from year-to-year.

As U.S. anthropogenic emissions of O_3 precursors decline, the relative importance of "U.S. background" to total surface O_3 rises. U.S. background O_3 is defined here as the O_3 levels that would exist in the absence of U.S. anthropogenic emissions of O_3 precursors, nitrogen oxide (NO_x) and non-methane volatile organic compound (NMVOC) precursors.). U.S. background O_3 thus includes naturally occurring O_3 as well as O_3 produced from global methane (including U.S. anthropogenic emissions) and from O_3 precursor emissions outside of the U.S.A. Jaffe et al. (2018) review the current understanding on U.S. background O_3 from models and observations, and its relevance to air quality standard setting and implementation. Previous studies estimating background O_3 over the United States found that background sources of O_3, including stratospheric O_3 intrusions (Lin et al., 2012, 2015a), increasing Asian anthropogenic emissions (Lin et al., 2015b), and more frequent wildfires in summer (Abatzoglou and Williams, 2016; Jaffe, 2011; Yang et al., 2015), may present challenges to obtaining the O_3 standard, especially since regional emission controls may be offset by a warming climate (Fiore et al., 2015). At high-altitude Westernsites in the western U.S.A.

(WUS) ~~sites~~ in spring, the influence from stratospheric intrusions and foreign transport, combined with relatively deep planetary boundary layers, can lead to high background O_3 events (Fiore et al., 2002; Zhang et al., 2011). Lin et al. (2017) investigated surface O_3 trends over the U.S.A. from 1980-2014 with the GFDL AM3 model and found that emissions controls decreased the 95th percentile summer O_3 values in the ~~Eastern~~eastern U.S.A. (EUS) by ~~0.2–0.4~~5-10 ppb ~~yr^{-1}~~ over 1988-2014, but rising Asian emissions ~~offset the effect of U.S. emissions reductions, leading to~~increased by 2-8 ppb ~~increases in monthly mean O_3~~ at individual sites in the WUS over the period (Lin et al., 2016).

Earlier work in the GEOS-Chem model analyzing background O_3 during a single meteorological year noted a tendency for the model to underestimate springtime O_3 at high-altitude WUS sites but overestimate summertime O_3 over the EUS ~~(e.g. Fiore et al., 2002, 2003; Wang et al., 2009; Zhang et al., 2011, 2014). Identifying the extent to which these biases reflect poor representation of U.S. anthropogenic versus background sources is relevant for assessing uncertainties in estimates of background O_3 on days when the O_3 NAAQS is exceeded. We build upon these prior studies by analyzing MDA8 O_3 measurements and 9-year model simulations spanning 2004-2012 from the GEOS-Chem 3D global chemistry-transport model (CTM). We use a suite of GEOS-Chem sensitivity simulations to estimate the influence from various individual background sources on O_3 concentrations and the interannual variability in background O_3 levels, with a focus on the highest 10 events in each EPA region during each summer (JJA) or year.~~(e.g. Fiore et al., 2002, 2003; Wang et al., 2009; Zhang et al., 2011, 2014). Identifying the extent to which these biases reflect poor representation of U.S. anthropogenic versus background O_3 sources is relevant for assessing uncertainties in estimates of background O_3 on days when the O_3 NAAQS is exceeded. We build upon prior studies by analyzing MDA8 O_3 measurements and 9-year model simulations spanning 2004-2012 from the GEOS-Chem 3D global chemistry-transport model (CTM). A suite of sensitivity simulations in which different emissions of O_3 precursors are perturbed allows us to identify which sources are contributing the most to observed high-O_3 days and on the days with the highest model bias. We assess here whether biases in the model reflect problems in the modeled transported background versus O_3 produced within the U.S. from both background and anthropogenic sources. In addition, the availability of these simulations for 2004-2012 allow us to investigate the year-to-year variability in background sources and the extent to which this variability is relevant for observed high events, and therefore, potentially to attaining the O_3 NAAQS. Though coarse resolution global models such as GEOS-Chem will mix emissions into the same grid cell that may remain separate in the real atmosphere, a global model is necessary to quantify background O_3 transported intercontinentally, including that produced via oxidation

of methane. We estimate the influence from various individual background sources on O_3 concentrations and the interannual variability in background O_3 levels with a focus on the highest 10 events in each of the 10 U.S. EPA regions during each summer (JJA) or year. We aim to answer the following questions: (1) Which sources exert the strongest influence on O_3 on the ten days with the highest model biases against observations? (2) Which background sources influence total O_3 the most on average versus the 10 highest O_3 days? (3) Which sources influence the interannual variability of O_3 in each region on average versus the 10 highest O_3 days?

2 Observations and model simulations

2.1 Observations

We use observed 2004-2012 MDA8 O_3 data from the EPA Air Quality System (AQS) network of urban, suburban, and rural monitoring sites, the Clean Air Status and Trends Network (CASTNet), and the Mount Bachelor Observatory (https://digital.lib.washington.edu/researchworks/browse?type=subject&value=Mt.+Bachelor+Observatory) in Oregon. MDA8 O_3 values for the AQS sites were ~~download~~downloaded from http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html#Daily (2004-2012 data last updated June 28, 2013). This dataset includes 1644 total sites from the contiguous U.S.A. from 2004-2012 with 1207 to 1333 sites collecting data each year (U.S. Environmental Protection Agency, 2014) (~~Supplemental Figure 1). The number of AQS sites measuring data per year is listed in Supplemental Figure 1.~~Supplemental Table 1).

The CASTNet (ftp://ftp.epa.gov/castnet/data) O_3 monitoring sites are located in rural areas away from emission sources and densely populated regions. CASTNet sites are designed to capture background O_3 levels and characterize ~~the~~ broad spatial and temporal O_3 trends~~ of air pollutants~~. We calculate the MDA8 O_3 concentration from hourly values at 108 CASTNet sites with data between 2004-2012, requiring at least 18 hours of data per day for each MDA8 O_3 calculation.

The Mount Bachelor Observatory, established in 2004 by the University of Washington Jaffe Research Group, is located 2.7 km above sea level on the summit of Mount Bachelor, an extinct volcano in the Cascade Mountains of central Oregon. It provides an estimate of baseline O_3 levels over the West Coast of the United States. Baseline O_3 is ~~tropospheric~~defined as the O_3 ~~concentrations~~concentration at sites ~~that have a~~with negligible influence from local emissions (National Research Council, 2010). ~~Here we take all hourly O_3 concentrations from Mount Bachelor and calculate the MDA8 O_3 concentrations for 2004-2012. Daily averages are included only if at least 18~~

~~hours of data are available per day. As we did not archive three-dimensional high frequency data, all MDA8 O_3 values from the model are sampled at the lowest surface layer for comparison to observation sites. Monthly MDA8 O_3 averages from Mount Bachelor are included only if at least 20 days out of the month contain valid daily data. For our comparison to monthly mean measurements at Mount Bachelor, we sample the model at the height closest to 2.7 km.~~

Baseline O_3 is a measurable quantity and differs from background O_3 in that it contains some influence from U.S. anthropogenic emissions that were not recently emitted but contributed to the global background. This station represents variations in baseline O_3 concentrations~~In order to evaluate the GEOS-Chem model O_3 simulation (described below in Sect. 2.3) at a spatial scale comparable to the coarse horizontal resolution global grid (2º x 2.5º), we use an available 1° x 1° grid of interpolated surface MDA8 O_3 measurements described by Schnell et al. (2014). We degrade the Schnell et al. (2014) dataset to 2° x 2.5° to match that of the GEOS-Chem simulations.~~

~~2.2~~2.1 ~~Analysis regions~~

~~Each observational site in the EPA AQS and CASTNet datasets is linked to one of the 10 EPA air quality regions (Figure 1) based on which state the site is in. The Mount Bachelor data were included with the Region 10 (Pacific Northwest) sites (Table 2) as a station representative of variations in baseline O_3 concentrations in the U.S.A.~~ (Baylon et al., 2016) and is analyzed as a standalone site in Section 3.2 given the relevance of high-altitude measurements for downwind surface O_3 (Stauffer et al., 2017). We take all hourly O_3 concentrations from Mount Bachelor and calculate the MDA8 O_3 concentrations for 2004-2012. Daily averages are included only if at least 18 hours of data are available and monthly averages require at least 20 days with valid 24-hour mean or MDA8 data. For our comparison to monthly average O_3 at Mount Bachelor Observatory, we sample the model both at the level closest to 2.7 km and at the surface.

We use temperature data from a 0.5º x 0.5º resolution gridded dataset developed by Fan and van den Dool (2008) using data from the Global Historical Climatology Network (GHCN) and the Climate Anomaly Monitoring System (CAMS). Each observational site is matched to the temperature model grid cell it falls in and the average monthly temperature is computed by averaging across all the sites in each region.

In order to evaluate the GEOS-Chem model O_3 simulation (described below in Sect. 0) at a spatial scale comparable to the coarse horizontal resolution global grid (2º x 2.5º), we use an available 1° x 1° grid of surface MDA8 O_3 measurements, interpolated from the AQS, CASTNet, and Canadian NAPS networks (Schnell and

Prather, 2017). We degrade this 1° x 1° dataset to 2° x 2.5° to match the horizontal resolution of the GEOS-Chem simulations. As we did not archive three-dimensional high frequency data, all MDA8 O_3 values from the model are sampled at the lowest surface layer for comparison to observational sites.

2.2 Analysis regions

~~even though it is not a regulatory monitor. Similar to~~Each observational site in the EPA AQS and CASTNet datasets is linked to one of the 10 U.S. EPA air quality regions (Supplemental Figure 1) based on which state the site is in. The Mount Bachelor data were included with Region 10 (Pacific Northwest) sites even though it is not a regulatory monitor. Following Reidmiller et al. (2009), we select two regions, the Southeast (Region 4) and Mountains and Plains (Region 8), as representative regions for the EUS and WUS for illustration purposes in the main text. Figures ~~showing our results~~ for the other ~~six~~eight regions are included in the supplement.

~~We~~To find the daily mean O_3 concentration within each region, we first match each observational site to the model grid within which it falls. We then average across all sites in each region to obtain a regional mean MDA8 O_3 value ~~for each day.~~in the observations and in the model. From the regionally averaged observed MDA8 O_3, we find: (1) the ten days with the highest observed O_3 during each year (hereafter, O_3_top10obs days~~;~~), similar to the definition for extreme events used in ~~Schnell et al., 2014)~~Schnell et al. (2014), (2) the ten days with the highest O_3 observations during each season (hereafter, O_3_top10obs_MAM, O_3_top10obs_JJA, and O_3_top10obs_SON), and (3) the 4th highest MDA8 O_3 within each year. In addition, we sample the model to find the ten days each year with the highest positive biases. There is at most a 2-6 day overlap between the top 10 O_3 Base days and the top 10 most biased days in 2004-2012 across all regions, but during most years, the overlap is around 0-2 days. We restrict our analysis to examining the top 10 observed O_3 days as these days are most relevant from a policy perspective. We use O_3_top10obs as our primary metric, however, instead of the policy-relevant 4th highest O_3 because the model bias is ~~comparatively~~typically lower~~; on the days with the 4th highest values, the model bias is generally more strongly negative in the west and South Central regions and more strongly positive in the Midwest than~~ on O_3_top10obs days (Supplemental Figure 2~~,~~ versus Supplemental Figure 3~~). In addition, while the model rarely captures the exact day of the 4th highest MDA8 O_3 event, there is a 3-4 day overlap on average between the O_3_top10obs days and the highest 10 MDA8 O_3 days in the model. This overlap is similar to the 3 and 6 day overlap Jaffe et al. (2017) found in their regional models for May 1st to September 29th, 2011.~~

). On the days when the 4th highest values occur, the model bias is generally more strongly negative in the west and South Central regions and more strongly positive in the Midwest than on O_3_top10obs days (Supplemental Figure 2, Supplemental Figure 3). In addition, while the model rarely captures the exact day of the 4th highest MDA8 O_3 event, there is a 3-4 day overlap on average between the O_3_top10obs days and the highest 10 MDA8 O_3 days in the model. This overlap is similar to the 3 and 6 day overlap Jaffe et al. (2018) found in their regional models for May 1st to September 29th, 2011.

2.3 GEOS-Chem model simulations

We use the GEOS-Chem v9_02 global 3D chemical transport model (CTM) (http://www.geos-chem.org) simulations driven by Modern-Era Retrospective analysis for Research and Applications (MERRA) reanalysis meteorology from the NASA Global Modeling and Assimilation Office for 2004-2012 (Rienecker et al., 2011). The MERRA reanalysis is available at 1/2º by 2/3º horizontal resolution, which we degrade here to 2º by 2.~~5° horizontal resolution.~~5º horizontal resolution. MERRA meteorology captures summer mean surface temperatures to within 1-2 K across U.S. regions and precipitation to within 0.5 mm d^{-1} except for over the Northern Great Plains where a positive bias exceeds 1 mm d^{-1}, but the variance in summer mean precipitation is lower than observed in some regions (Bosilovich, 2013). While interannual variability in cloudiness observed at weather stations is largely captured by MERRA, the reanalysis generally underestimates cloud cover and thus overestimates observed downward surface shortwave fluxes (Free et al., 2016). Methane surface concentrations are prescribed each month using spatially interpolated surface distributions from NOAA Global Monitoring Division flash data. We use the standard v9_02 chemical mechanism which includes recycling of isoprene nitrates (Mao et al., 2013) in contrast to the mechanisms used in earlier versions of GEOS-Chem (e.g., Zhang et al., 2014 as discussed in Fiore et al., 2014). Anthropogenic base emissions are from the Emission Database for Global Atmospheric Research (EDGAR) version 3.2-FT2000 inventory (Olivier et al., 2005) for inorganic compounds and the REanalysis of the TROpospheric chemical composition (RETRO) inventory (Hu et al., 2015; Schultz, 2007) for organic compounds. Inorganic emissions are overwritten by regional inventories for the U.S. (EPA National Emissions Inventory 2005), Canada (Criteria Air Contaminants), Mexico (Big Bend Regional Aerosol and Visibility Observational study; Kuhns and Green, 2003), Europe (European Monitoring and Evaluation Programme; Auvray and Bey, 2005), and South and East Asia (Streets et al., 2006). Separate global inventories are used for ammonia (Bouwman et al., 1997), black carbon (Bond et al., 2007; Leibensperger et al., 2012), and ethane (Xiao et al., 2008). Anthropogenic surface emissions have diurnal and

monthly variability, some with additional weekly cycles, and are scaled each year on the basis of economic data and estimates provided by individual countries, where available ~~(Van Donkelaar et al., 2008). Aircraft emissions are from the Avian~~(van Donkelaar et al., 2008). The model does not include daily variations in U.S. anthropogenic emissions associated with higher electricity demand on hotter days (e.g., Abel et al., 2017). Aircraft emissions are from the Aviation Emissions Inventory Code (AEIC) inventory (Stettler et al., 2011) and shipping emissions are from International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Lee et al., 2011; Wang et al., 2008). Biomass burning emissions follow the interannually-varying monthly Global Fire Emissions Database version 3 (GFED3) inventory driven by satellite observations of fire activity (Giglio et al., 2010; Van Der Werf et al., 2010). Biofuel emissions are constant (Yevich and Logan, 2003). Biogenic VOC emissions from terrestrial plants follow the Model of Emissions of Gases and Aerosols from Nature (MEGAN) scheme version 2.1 (Guenther et al., 2012) and vary with meteorology (Barkley et al., 2011). ~~Emissions of NO_x from soil microbial activity follow Hudman et al. (2012). Methane surface concentrations are prescribed each month using spatially interpolated surface distributions from NOAA Global Monitoring Division flash data.~~Global and U.S. emissions are 29.5 Tg N yr^{-1} and 5.2 Tg N yr^{-1} respectively, for anthropogenic NO_x emissions (including biofuels); 4.2 Tg N yr^{-1} and 0.1 Tg N yr^{-1} for biomass burning; 8.7 Tg N yr^{-1} and 0.9 Tg N yr^{-1} for soil NOx; 6.7 Tg N yr^{-1} and 1.0 Tg N yr^{-1} for lightning NOx; 466.1 Tg C yr^{-1} and 20.6 Tg C yr^{-1} for isoprene emissions. Emissions for NO_x sources and isoprene are provided globally and within the U.S.A. for each year in Supplemental Table 3.

We first perform a base simulation (O_3_Base) ~~in which~~with all emissions ~~are prescribed normally~~turned on for 2003-2012. We ~~then perform~~conduct a parallel suite of sensitivity simulations, in which selected sources are removed. In all simulations, we discard 2003 from our analysis as initialization. Our first set of sensitivity simulations estimates three different "background" definitions: (1) "North American Background" (denoted O_3_NAB) in which ~~we remove individual sources (Table 1), including (1) U.S.~~ anthropogenic emissions within Canada, Mexico, and the U.S.A. are set to zero, but ~~maintaining~~methane surface abundances are kept at present-day ~~methane concentrations; the O_3 in this simulation provides an estimate of~~ values; (2) "U.S. background ~~O_3 (hereafter,"~~ (O_3_USB), which is similar to O_3_~~USB); (2) an otherwise identical simulation that also excludes those~~NAB except only U.S. anthropogenic emissions ~~from Mexico and Canada, O_3 in this simulation is referred to as "North American Background" O_3 (O_3_NAB);~~are set to zero; (3) ~~wildfire emissions, (4) biogenic VOC emissions, (5) Soil NO_x, and (6) Lightning NO_x. In addition, we perform a "natural" simulation~~"Natural background" (O_3_NAT), in which all

anthropogenic emissions have been removed globally and methane is prescribed at preindustrial levels ~~to provide an~~. We estimate ~~of "natural"~~ Canadian and Mexican influence (O_3 ~~(~~_CA+MX) on U.S. O_3 by subtracting O_3_NAB from O_3_USB; the influence from intercontinental pollution transport plus global methane (O_3_ICT+CH_4) is estimated by subtracting O_3_NAT~~). In all~~ from O_3_NAB. A second set of sensitivity simulations~~, we discard 2003 from our analysis as initialization. We~~ enable us to estimate the contribution of ~~each~~ individual ~~sector~~"background" sources to ~~the~~ total ~~concentration~~ simulated surface O_3 by subtracting ~~the O_3 in each sensitivity~~a simulation ~~in which~~with that source ~~has been removed~~shut off from the O_3_Base simulation~~. As in all "zero-out" perturbation simulations,~~: (1) O_3_NALNO_x by turning off North American lightning NOx; (2) O_3_SNO_x by zeroing out global soil NOx; (3) O_3_BVOC by zeroing out terrestrial biogenic VOC emissions (we also examine this "O_3_noBVOC" simulation in Section 3.3); (4) O_3_BB by zeroing-out biomass burning emissions, as summarized in Table 1. Due to non-linearities in atmospheric photochemistry ~~make this a simple estimate~~, these "zero-out" estimates of ~~the contribution of each~~ source~~, and the contribution of each source depends~~ contributions depend on the presence of all other precursor emissions at present-day levels (e.g., the impact of BVOCs emissions is sensitive to the amount of anthropogenic NO_x emissions~~). Hereafter, the terms listed in the "Notation" column of Table 1 will be used to refer to the influence of each source on total O_3 (O_3_Base). Note that this~~ in the Base simulation). This set of model simulations does not directly isolate stratospheric O_3 or Asian influences. Previous work has shown that stratospheric O_3 can increase springtime O_3 levels by 17-40 ppb in the WUS ~~in spring~~ when MDA8 O_3 levels are 70-85 ppb, and Asian emissions can contribute 8-15 ppb to MDA8 O_3 on days above 60 ppb (Lin et al., 2012, 2015a). Stratospheric and Asian influences are included ~~as part of our~~in O_3_USB ~~estimates~~; Asian influences are included in O_3_ICT+CH_4~~,~~; and O_3_NAT ~~contains the influence of~~includes stratospheric O_3 ~~along with natural~~ , biogenic ~~precursor~~ emissions of O_3 precursors, wildfires, and lightning NO_x. As O_3_BVOC includes O_3 produced from biogenic VOC reacting with both natural and anthropogenic NO_x, O_3_USA and O_3_BVOC are not additive. O_3_BVOC thus contributes to both O_3_USA and O_3_USB.

3 Model evaluation

3.1 MDA8 O_3 distributions

~~Previous studies have found that averaging all observational sites within a model grid cell tend to disproportionately represent urban stations, especially when looking at high O_3 days (Schnell et al., 2014).~~ To evaluate the ability of our coarse resolution model to capture observed high-O_3 events, we compare the MDA8 O_3

~~simulated by GEOS-Chem to the observations in two ways. First, we use the Schnell et al. (2014) gridded dataset degraded to the model resolution. Second, we compare each individual observational site to the model grid cell within which it resides.~~averaged over each of the 10 EPA regions simulated by GEOS-Chem to the observations in two ways. In the first method, we use the (Schnell and Prather, 2017) gridded dataset degraded to the model resolution and sample the model directly at each of the degraded Schnell grid cells prior to calculating the regional average. In the second method, we sample the model grid cell containing each individual observational site (EPA AQS, CASTNet, and Mount Bachelor Observatory) prior to calculating the regional average. The model is biased positively with either method (Figure 1a, b), but the shape of the model distribution constructed with the latter approach (Figure 1b) better matches the observed distribution than that of the former (Figure 1a). Matching individual sites to the nearest model grid (Figure 1b) ~~results in~~yields a better estimate of high-O_3 days; the model overestimates the percentage of days above 70 ppb by about three times when we match to individual measurement sites (3.14% of days are above 70 ppb in the observations versus 9.92% in model) but by about ten times in comparison to the re-gridded Schnell (2014) dataset (0.37% of days are above 70 ppb in the observations versus 3.91% in the re-gridded dataset).

Simulated seasonal mean MDA8 averaged over the full 2004-2012 period is higher than observed by 5-30 ppb (Figure 2a, b, c), with the largest biases typically occurring in the Northeast and Midwest. The model bias is highest in summer (JJA) (15-30 ppb at most sites), followed by fall (SON) (10-20 ppb) (Figure 2a, b, c). Recent work in a newer version of GEOS-Chem attributes some of the positive model bias in the EUS to excessive NO_x emissions in the 2011 National Emission Inventory (NEI) (Travis et al., 2016), an inability of the model to resolve vertical mixing in the boundary layer, and a weak response to cloud cover (Travis et al., 2017). Travis et al. (2016) find that the 3.5 Tg N y^{-1} NEI 2011 estimate for U.S. fuel NOx emissions is too high and contributes to excessive surface O_3. Our simulations include even higher U.S. fuel NOx emissions of 4.4 Tg N y^{-1} during 2010-2012 (Supplemental Table 3), implying that some portion of the model O_3bias reflects excessively high anthropogenic NOx emissions (Travis et al., 2016). The low bias in cloud cover in the MERRA meteorology and associated overestimate in downward shortwave surface radiation (Free et al., 2016) may also contribute to excessive O_3

production in the model. The model is closest to the observations in spring, with a positive bias usually <10 ppb over the eastern states and generally within ±5 ppb over most western sites (Figure 2a, b, c).

3.2 Baseline O_3 at Mount Bachelor

~~3.23.1 Baseline O_3 at Mount Bachelor~~

Mount Bachelor Observatory ~~is located at 2.7 km above sea level, where it~~(MBO) regularly samples free tropospheric O_3 and is rarely influenced by local anthropogenic emissions (Reidmiller et al., 2009). It is therefore, a valuable site for examining baseline O_3 ~~values. Figure 3 compares modeled and observed monthly mean O_3 at Mount Bachelor. The observations peak in springtime and then fall in the summer months. The model, however, has a maximum in and underestimates springtime baseline O_3. We infer, consistent with our analysis below, that the model does not resolve springtime high O_3 events, possibly reflecting an underestimate of stratospheric influences (see Fiore et al., 2014; Zhang et al., 2011; 2014). The model indicates that O_3_USB dominates O_3_Base (Figure 3). Even at this baseline site, however, the model indicates that U.S. anthropogenic emissions enhance monthly mean O_3 by at least a few ppb (estimated as the difference between O_3_Base and O_3_USB).~~ . In Supplemental Figure 4, we compare the observed 24-hour and MDA8 O_3 concentrations at MBO for 2004-2012. The observed O_3 concentrations vary from year to year, and by definition, MDA8 O_3 is a few ppb higher than the 24-hour mean mixing ratios. However, the seasonal pattern is similar across both metrics, with a springtime peak, maximum in April, and a secondary summertime peak in July.

Figure 3 compares modeled and observed monthly mean 24-hour O_3 concentrations at the grid box that contains Mount Bachelor. For the model, we examine O_3_Base and O_3_USB 24-hour average concentrations at 2.7 km, the height of the Mount Bachelor Observatory, as well as at the surface. It is important to note that the diurnal variations on the mountain may not be well captured by the CTM, due to upslope (daytime)/downslope (nighttime) flow. We focus on the 24-hour average because we only archived hourly O_3 fields from the model at the surface and thus, do not have the MDA8 O_3 metric available at 2.7 km. The year-to-year variability is smaller in the model than observed (narrower shaded range). In all months, the O_3_Base and O_3_USB values are higher by 9-14 ppb and 11-21 ppb, respectively, at 2.7 km than at the surface. The model captures the magnitude of the observed springtime peak at 2.7 km, but summertime values are too high, with an overall peak in August. O_3_USB contributes a greater fraction to O_3_Base at 2.7 km (92-94%) than at the surface (72-94%). The simulated seasonal cycle differs at the

surface, peaking in spring (March-April) and in September. In 2012, the observations show equivalent springtime and summertime peaks, more similar to the modeled seasonal cycle. While the observations generally decline from spring into summer, the model indicates an increase, leading to a substantial model overestimate during summer in most years. This model bias occurs across much of the U.S.A. as we show below.

Our sensitivity simulations enable us to interpret the sources contributing to the simulated seasonal distribution. The model indicates that at MBO, O_3_USB is the major component of O_3_Base, including during the summertime overestimate. In turn, the model indicates that the seasonality of O_3_USB is largely driven by O_3_NAT, which includes the influence from biogenic VOC and NOx, lightning NO_x, as well as stratospheric O_3. O_3_ICT+CH_4 contributes around 15 ppb at 2.7km and 5-10 ppb at the surface (Figure 3). The model does suggest a springtime peak influence from O_3_ICT+CH_4 in the WUS, consistent with earlier work (e.g., Task Force on Hemispheric Transport of Air Pollution, 2010). Even at this baseline site, the model indicates that O_3_USA enhances monthly mean O_3 by at least a few ppb at 2.7 km; at the surface, the model simulates a seasonal cycle for O_3_USA that is typical of photochemical production from regional precursor emissions. O_3_CA+MX is less than a few ppb at MBO whether the model is sampled at 2.7 km or the surface (not shown).

3.3 Magnitude and timing of high-O_3 events

~~Simulated seasonal mean MDA8 averaged over the full 2004-2012 period is higher than observed by 5-30 ppb (Figure 4a, b, c), with the largest biases typically occurring in the Northeast and Midwest. The model bias is highest in summer (JJA) (15-30 ppb at most sites), followed by fall (SON) (10-20 ppb) (Figure 4a, b, c). Recent work in a newer version of GEOS-Chem attributes some of the positive model bias in the EUS to excessive NO_x emissions in the 2011 National Emission Inventory (NEI) (Travis et al., 2016), an inability of the model to resolve vertical mixing in the boundary layer, and a weak response to cloud cover (Travis et al., 2017). The model is closest to the observations in spring, with a positive bias usually <10 ppb over the eastern states and generally within ±5 ppb over most western sites (Figure 4a, b, c).~~ On O_3_top10obs days, ~~however,~~the model biases are typically lower than on average days (Figure 2~~, Table 3;~~, Table 2; see also year-by-year maps in Supplemental Figure 2). At some WUS sites, the model underestimates O_3 levels during the highest events by 10-20 ppb. ~~We note that the~~The model systematically underestimates O_3 in the Central Valley of California in all three seasons, which we attribute to the

inability of the coarse model resolution to resolve topographical gradients and valley circulations (or stagnation) in this region which experiences some of the highest observed O_3 in the nation.

We compare the MDA8 O_3 distributions in the observations versus the model (O_3_Base) during the 10 most biased days in each of the ten regions across the nine years (900 total events). These "most-biased" days in the model tend to fall around the observed median (Figure 1c) during the warm season (June - October), with almost 40% of the days falling in August alone (~~Figure 5~~Supplemental Figure 5), and are 9-45 ppb higher than the observations (circles in ~~Figure 6, Supplemental Figure 3~~Supplemental Figure 6). We analyze the perturbation simulations (Table 1) to identify which sources influence simulated O_3 most strongly on the "most-biased" days versus on average (i.e., all 365 or 366 days), which we assume are also likely the main drivers of the bias. In all regions, the largest sources on the "most-biased" model days are O_3_USA (3-30 ppb higher MDA8 O_3 than on average with the exception of the Pacific SW where O_3_USA is smaller than on average days), O_3_BVOC (by 1-15 ppb), and O_3_SNO_x (by 1-10 ppb; Figure 4~~, Supplemental Figure 3).~~, Supplemental Figure 6). By contrast, O_3_ICT+CH_4 is up to a few ppb higher on average days than on the most-biased model days. ~~The 10 most biased days in the model tend to be 10°C warmer than average (Figure 6, Supplemental Figure 3), contributing to the higher O_3_BVOC and O_3_SNO_x. We emphasize that O_3_USA and O_3_BVOC are not additive as anthropogenic NO_x reacts with biogenic VOC to produce O_3.~~

To explore possible drivers of model biases across the different seasons, we evaluate the timing of the highest ten events across each year in the O_3_Base, O_3_USB, and O_3_noBVOC (BVOCs shut off) simulations for each region (900 events). We bin these 900 events by month and calculate the percentage of the total events that fall within each month. Note that all the top ten days fall between March and October. The standard model (O_3_Base) underestimates the occurrence of high events early in the O_3 season (March-June) and overestimates them later in the season (July-September) (~~Figure 7). While the model indicates that most top ten O_3 days fall between July-August (35% each), the observations show that May through August each contain around 15-25% with the maximum in June at 25%. When we examine the highest ten O_3 events in the O_3_USB case (U.S. anthropogenic emissions shut off), we see 5-10% fewer top-ten events in July and August (27% in July and 28% in August), suggesting that O_3_USA is contributing most to the temporal shift (and general summertime overestimate) relative to the observations. The O_3_USB case does capture some early spring events in April (5%) and May (10%), though still fewer than observed (12% and 17% respectively). In the O_3_noBVOC case, there are~~

~~5-10% more events during April and May than in the O_3_Base case, but the shortage of high spring O_3 events remains. The lack of high events in spring may stem from the springtime underestimate in this model, particularly at high altitude sites (e.g., Figure 5; see also Figures 4 and 6 of Fiore et al. (2014)), and may reflect poor representation of stratospheric O_3 intrusions at the coarse resolution of the CTM (Zhang et al., 2014). The summertime overestimate of high O_3 events is less pronounced in the O_3_noBVOC case than in the O_3_Base case, implying that BVOCs are also contributing to the misplaced seasonal timing of the highest events, either through excessive O_3 production or a missing coincident sink.~~Supplemental Figure 7). While the model indicates that most top ten O_3 days fall between July-August (35% each), the observations show that May through August each contain around 15-25% with the maximum in June at 25%. Both O_3_noBVOC and O_3_USB shift the relative timing of the 10 highest O_3 events towards April and May compared to O_3_Base, but the shortage of high springtime O_3 events remains (Supplemental Figure 7). The lack of high events in spring may reflect in part poor representation of stratospheric O_3 intrusions at the coarse resolution of the CTM (Lin et al., 2012; Zhang et al., 2014), in addition to the role of U.S. anthropogenic and BVOC emissions in the temporal mis-match as indicated by the improvements to the timing that occur in the O_3_USB (U.S. anthropogenic emissions shut off) and O_3_noBVOC simulations. In addition to contributions from these sources, poor representation of O_3 sinks may contribute to the model biases. For example, Makar et al. (2017) suggest that failing to represent canopy turbulence and shading effects on photolysis can lead to high-O_3 biases in models.

3.4 Interannual variability

~~Figure 8~~Supplemental Figure 8 shows the Pearson correlations coefficients (r) between monthly average observed and O_3_Base values from 2004-2012. In May, correlations are generally strong ($r \geq 0.9$) in the Mid-Atlantic and Southeast regions, but much lower (r = 0.2) in the New England region. This pattern may reflect shortcomings in representing the onset of BVOC emissions. In July, the regions flip, with lower correlations in the Southeast and higher correlations in New England. At some sites in the WUS, lower correlations occur during summer months, which may be tied to excessive influence from lightning NO_x advected from Mexico (see also Zhang et al., 2011; 2014) or anomalous events such as wildfires that are not well- captured by the model.

In general, correlations only average about r = 0.2 in the winter and early spring over much of the United States (Supplemental Figure 8); the drivers for these weak correlations may be connected to the model tendency to

underestimate the occurrence of springtime high-O_3 events. From May to September, however, the months during which high-O_3 events are most likely to occur, the correlation between 2004-2012 observed and simulated O_3 monthly averages over much of the contiguous United States exceed r = 0.7 (~~Figure 8, Supplemental Figure 4).~~Supplemental Figure 8). We conclude that the model broadly captures monthly variations from year-to-year during the warm season and can thus be applied to interpret the role of background sources in contributing to interannual variations during most of the high-O_3 season. We note that Clifton et al. (2017) found that the GEOS-Chem model does not capture interannual variability in deposition velocities observed at Harvard Forest, MA, but it is unclear to what extent this process would amplify or dampen interannual variability associated with changes in emissions.

4 Influence of individual sources on average versus high-O_3 days

In ~~Tables 3~~Table 2 and ~~4,~~Table 3, we report the influence of the O_3 sources defined in ~~Table 1~~Table 1 on average versus O_3_top10obs days separately for spring (MAM), summer (JJA), and fall (SON) (ten days from each of the nine simulation years for 900 events for each region and season). We also report the difference in source influences between average and O_3_top10obs days, which we interpret as the enhancement from that source relative to average conditions.

We first consider the average ranges in MDA8 O_3 contributed by the various sources. Both O_3_USA and O_3_USB tend to follow the seasonal cycle of O_3_Base, with highest abundances in summer. The model indicates that O_3_USB is 30-~~60~~50 ppb (range over regions) during summer and highest over the WUS. O_3_USA is generally 20-30 ppb over the EUS in summer, but only 10-20 ppb over the WUS~~.~~ (Table 2). O_3_ICT+CH_4 averages 2-13 ppb over all regions and is highest in spring (8-13 ppb compared to 2-11 ppb in summer and 6-12 ppb in fall) ~~(Table 4,~~(Table 3, Figure 5, Supplemental Figure 9). O_3_NALNO$_x$ has a relatively minor influence (at most 1.5 ppb) in all regions and seasons. The influence from O_3_CA+MX is generally less than a couple of ppb~~.~~ except in NY+NJ and New England where it can be as much as 4-7 ppb (Table 3, Supplemental Figure 9).

We interpret the "difference" lines in ~~Tables 3~~Table 2 and ~~4~~Table 3 as the enhancements from each source on high days in each season (O_3_top10obs_MAM, O_3_top10obs_JJA, O_3_top10obs_SON) relative to average conditions. Over all regions, O_3_BVOC and O_3_SNO$_x$ influence O_3_Base more on O_3_top10obs days (for all seasons) than on average ~~days~~ whereas O_3_ICT+CH_4 is typically lower by up to 3 ppb on O_3_top10obs days (for all

seasons) than on average days ~~(Table 3, Table 4,~~(Table 2, Table 3, Figure 5, Supplemental Figure 9). O_3_USA is 8-11 ppb higher on O_3_top10obs_JJA days versus average ~~days~~ over the New England, NY+NJ, Mid-Atlantic, Midwest, and South Central regions, but only up to 5 ppb higher over other regions ~~(Table 3, Figure 9, Supplemental Figure 5).~~(Table 2, Figure 5, Supplemental Figure 9). The model indicates an even stronger anthropogenic enhancement (up to 19 ppb) on O_3_top10obs_SON days in some EUS regions ~~(Table 3).~~(Table 2). O_3_USB is enhanced on O_3_top10obs_JJA days by 2-12 ppb relative to average ~~days~~, with the smallest enhancements occurring in the Mid-Atlantic, Southeast, and Midwest regions, and the largest enhancements occurring in the Pacific NW. In contrast to all the other regions, O_3_USB is the dominant source enhancing O_3_top10days_JJA over the Mountains and Plains, Pacific NW, and Pacific SW regions (4-12 ppb for O_3_USB but < 5 ppb from either O_3_USA or O_3_BVOC). In line with earlier work reviewed by Jaffe et al. ~~(2017)~~(2017), enhanced O_3_USA dominates O_3_top10obs_JJA days over much of the U.S.A., whereas in the WUS, O_3_USB enhancements exceed O_3_USA enhancements on O_3_top10days_JJA. O_3_BVOC enhances O_3_top10obs days (for all seasons) by up to 9 ppb, with the influence often largest in fall (when O_3 formation is more sensitive to VOC; e.g., ~~Jacob et al., 1995~~Jacob et al., 1995). We re-emphasize that BVOCs contribute both to O_3_USA when reacting with anthropogenic NO_x and to O_3_USB when reacting with all other NO_x sources. In contrast to the sources discussed above, O_3_ICT+CH_4 influences average days by up to a few ppb more than on O_3_top10obs days (for all seasons), with the largest differences between average and high days occurring in EUS regions (1-3 ppb lower on O_3_top10obs days (for all seasons) in New England, NY+NJ, Mid-Atlantic; ~~Table 4, Figure 9~~Table 3, Figure 5, Supplemental Figure 9). O_3_NALNO_x is at most 2 ppb higher than average on O_3_top10obs days. The O_3_CA+MX influence is roughly equivalent (generally to within a ppb) on average versus O_3_top10obs days during all seasons ~~(Table 4).~~.

5 Interannual variability in the sources influencing high vs. average ground-level O_3

Despite its high mean bias and seasonal phase shift, the model does capture some of the observed interannual variability in observed O_3_top10obs_JJA MDA8 O_3 concentrations (~~Figure 8~~Figure 6, Supplemental Figure 10; r = 0.5 to $\geq$ 0.9). Comparing the 2004-2006 period with 2010-2012, both observed and simulated MDA8 O_3 concentrations on O_3_top10obs_JJA days hold steady or decrease across all regions. This change reflects opposing influences in the model: rising O_3_USB (by 2 ppb averaged over all regions) and declining O_3_USA concentrations (by 6 ppb averaged over all regions) (Figure 6~~, Table 5,~~, Table 4, Supplemental Figure 10). We note

that over the Pacific NW there is a 4 ppb decrease in O_3_USB from 2004-2006 to 2010-2012. Over this period, temperatures generally warm over the EUS, but slightly cool in the WUS. Within the ten regions, the model captures the sign of the changes in MDA8 O_3 over this period but not the magnitude (~~Table 5). We emphasize that~~Table 4). The model monthly mean temperatures in the model (from the MERRA reanalysis) closely match the observed GHCN+CAMS dataset (Supplemental Table 4). Table 4 shows that regions with O_3_USB increases generally experienced rising temperatures over this period, as the 2010-2012 period includes two of the warmest years on record~~(~~. Figure 6~~). Figure 10~~ shows that O_3_NAT tracks with O_3_USB~~, indicating that the year-to-year variability in O_3_USB is primarily driven in the model by meteorology as opposed to variability in upwind international anthropogenic emissions. O_3_USB~~ and ~~O_3_NAT on O_3_top10obs_JJA days generally track meteorological changes, with~~ temperature (dips in MDA8 O_3 ~~occurring~~occur during years with cooler temperatures (2008-2009) and increases in years with warmer temperatures (2011-2012~~) (~~), indicating that year-to-year variability in O_3_USB on O_3_top10obs_JJA days is primarily driven in the model by natural sources sensitive to meteorology rather than international O_3 transport (Figure 6, Supplemental Figure 10). ~~Note that although~~Although 2012 was the hottest year on average between 2004-2012 (except in the Pacific NW where 2004 was warmer by ~~around~~about a degree), it was not the hottest summer in all regions.

We find that O_3_USB drives the interannual variability on O_3_top10obs_JJA days in the WUS (r = 0.72-0.85 for O_3_USB versus O_3_Base, whereas r = 0.05-0.64 for O_3_USA versus O_3_Base; Supplemental Table 5). In NY+NJ, the Southeast, Midwest, South Central, and Plains regions, O_3_USB and O_3_USA both contribute to the interannual variability on O_3_top10obs_JJA days (r = 0.5-0.8 for both O_3_USB and O_3_USA versus O_3_Base) while in New England and the Mid-Atlantic regions, O_3_USA drives the interannual variability more than O_3_USB (r = 0.64 and 0.72 for O_3_USA versus O_3_Base but only 0.28 and 0.54, respectively, for O_3_USB versus O_3_Base; Supplemental Table 5).

Year-to-year variations in monthly average O_3_USB are relatively large, with 10-15 ppb differences between the highest and lowest O_3_USB years during the warmest months (Figure 7~~, Supplemental Figure 8).~~, Supplemental Figure 11). Seasonal variations also differ by region, especially during summer. For example, the western U.S. regions have a smooth seasonal cycle with O_3_USB concentrations rising from January to a peak in July and August, and then declining again. Interannual and seasonal variability in O_3_USB are generally greater in

the Southeast than in the Mountains and Plains, and Plains regions (Figure 7~~, Supplemental Figure 8).~~, Supplemental Figure 11). Year-to-year variability in O_3_BVOC is smaller than O_3_USB, with a maximum range of about 10 ppb between the highest and lowest years during August~~.~~ (Figure 7, Supplemental Figure 12). O_3_SNOx ranges by a few ppb throughout the summer in the Southeast, and by up to 6 ppb over the Mountains and Plains in August (Figure 7~~).~~, Supplemental Figure 13).

O_3_USA anomalies relative to the 2004-2012 average illustrate declining influence in all regions, with negative anomalies after 2007 on both O_3_top10obs and average days (Figure 8~~, Supplemental Figure 5). This finding is well established by earlier work demonstrating decreases in high-O_3 concentrations as a result of regional NO_x emissions reductions over the past few decades (Cooper et al., 2012, 2014a; Jaffe et al., 2017; Young et al., 2017). O_3_BVOC is the main driver of the high and low O_3 anomalies (up to ±5 ppb on O_3_top10obs_JJA days) from year-to-year.~~, Supplemental Figure 14). This finding is well established by earlier work demonstrating decreases in high-O_3 concentrations as a result of regional NO_x emissions reductions over the past few decades (Cooper et al., 2012, 2014a; Jaffe et al., 2018; Young et al., 2017). O_3_BVOC is the main driver of the high and low O_3 anomalies (up to ±5 ppb on O_3_top10obs_JJA days) from year-to-year (Figure 8, Supplemental Figure 15).

Specific events can affect O_3 in any given year. For example, in 2008, there were extensive fires across much of California in May, June, and July. In 2008, the Pacific SW region that includes California, Nevada, and Arizona, shows a positive anomaly in O_3_BB (> 1 ppb) on the O_3_top10obs days, stronger than during any other year in that region (Supplemental Figure 15). If we restrict our analysis solely to Reno, NV, the anomaly for O_3_BB was 7 ppb in July 2008 relative to the 2004-2012 July average (not shown). We emphasize that a single location can be more strongly influenced by a specific source than the regional averages on which we have focused.

Currently, the U.S. EPA uses a 3-year averaging period~~.~~ of the 4th-highest MDA8 O_3 to assess compliance with the O_3 NAAQS. We evaluate ~~here~~ the extent to which ~~the~~this 3-year averaging period removes interannual variability in meteorology (the grounds for the averaging~~). In~~) (Figure 9~~(~~, Supplemental Figure 16). The observed range is generally much smaller than the model estimate. We find that the 3-year average of the 4th highest day decreases the range by 2-6 ppb and ~~in Supplemental Figure 12, Supplemental Figure 13),~~5-18 ppb in the observations and O_3_Base respectively when compared to taking the 4th highest day in any given year when we ~~examine the range for each region~~look across all regions (Table 5). However, the 3-year average of the 4th highest

day still ranges from 3-9 ppb and 2-11 ppb in the observations and O_3_Base, respectively, across all regions (compared to 5-15 ppb and 10-36 ppb in the observations and O_3_Base on the 4th highest day in each individual year). Thus, while averaging across the years decreases the spread, variability remains. In keeping with our previous analysis of the O_3_top10obs days ~~between 2004-2012 in~~, we compare the ~~observations, O_3_Base, and O_3_USB. The dots indicate where~~spread of the 4th~~-~~ highest ~~MDA8~~ O_3 day ~~fell for~~in each ~~simulation. For the 2004 to 2012 period, the range~~ of the three~~-year averages of the observations is a few ppb lower than the annual~~ years to the range ~~covered by the 10 highest events (Figure 13, Supplemental Figure 12, Supplemental Figure 13). The annual range in the model (O_3_Base) sampled on O_3_top10obs days tends to be wider than the observed range (except for a few years in New England and NY+NJ) by as little as a few ppb to as much as 20 ppb. This modeled range overestimate lessens when averaged over three years (Figure 13a, b versus Figure 13c, d). We also include in Figure 12 (and Supplemental Figures 12 and 13) the range of the O_3_USB on~~of the O_3_top10obs days~~. While the three-year averaging period reduces the range in O_3_USB on the highest days, variability remains, and over the Mountains and Plains regions this~~ across each three year span; the 4th highest days can range almost as widely as the O_3_top10obs days in some years, but in other years, are clustered closer together (Figure 9). Figure 9 shows that the range in O_3_top10obs days for O_3_Base generally correlates with O_3_UBS in the WUS, suggesting that O_3_USB is the dominant ~~source influencing these~~influence on the high days ~~(Figure 13b, d).~~there, but there is little correlation in the EUS. We conclude that a three-year smoothing period is not long enough to eliminate entirely the interannual variability in ~~background~~ MDA8 O_3 levels~~.~~, and in the WUS, this interannual variability tends to reflect variations in O_3_USB.

6 Discussion and Conclusions

As air quality controls decrease U.S. anthropogenic precursor emissions to O_3, the relative importance of the background influence on total surface O_3 increases. We use O_3 MDA8 concentrations spanning 2004-2012 from the EPA AQS, CASTNet, and Mount Bachelor Observatory sites, and ~~various~~ sensitivity simulations from the global GEOS-Chem 3D chemistry transport model to estimate the influence from various individual background sources on O_3 in each of the ten EPA regions in the continental U.S.A. ~~We examine differences between background and U.S. anthropogenic influences on average and high O_3 days and on interannual variability.~~ The global scale of the GEOS-Chem model allows us to quantify intercontinental transport (including global methane) in addition to regional natural

and anthropogenic sources of O_3. The sensitivity simulations span nine years, allowing us to examine the role of these sources in contributing to interannual variability. Our analysis contrasts average- and high-O_3 days.

Correlations between monthly averages across 2004-2012 show that the model captures monthly variations from year-to-year, especially during summer (JJA). The model shows substantial variability in simulated U.S. background O_3 concentrations from year-to-year, on the order of 10-20 ppb between 2004-2012 in summer (Figure 7). We find that the extent to which the current three-year averaging period for assessing compliance with the National Ambient Air Quality Standard for O_3 succeeds in smoothing out interannual variability depends on the range in consecutive years, and thus varies by region and time period, but is generally not long enough to completely eliminate the interannual variability in background O_3 (Figure 9).

We find substantial biases in the severity (+0-19 ppb in maximum daily 8-hour average (MDA8) O_3) and timing of high-O_3 events in the model. The model underestimates the frequency of high events in spring, possibly associated with stratospheric intrusions (Fiore et al., 2014; Zhang et al., 2011; 2014). Future efforts would benefit from quantifying the stratospheric (as well as Asian) influence alongside the other background sources we consider. We find a stronger influence of U.S. anthropogenic emissions on regionally averaged MDA8 O_3 (up to 30 ppb) from BVOCs (up to 15 ppb) and soil NO_x (up to 10 ppb) on the ten most biased days as compared to average days. We conclude that regional production of O_3 is driving the pervasive high positive model bias in summer, as opposed to transported background, although our sensitivity simulations do not allow us to rule out the possibility of a coincident missing sink.

Our finding that BVOC emissions contribute to the summertime surface O_3 biases could reflect poor representation of the emissions (and subsequent oxidation chemistry). Earlier work has noted that MEGAN BVOC emissions are too high over California (Bash et al., 2016), Southeast Texas (Kota et al., 2015), the Ozarks in southern Missouri (Carlton and Baker, 2011), and across much of the U.S.A. (Wang et al., 2017). One recent model study uniformly reduced MEGAN isoprene emissions by 20% (Li et al., ACP 2018), but we did not apply any such scaling here. In regions that are highly NOx-sensitive, additional isoprene should not strongly influence O_3, as found over southeast Texas (Kota et al., 2015). While not eliminated entirely, the summertime model bias does lessen in the simulation with BVOC emissions set to zero, suggesting that the O_3 bias is indeed exacerbated if BVOC emissions are overestimated in the model.

~~We find substantial biases in the severity (+0-19 ppb in maximum daily 8-hour average (MDA8) O_3) and timing of high-O_3 events in the model. The model underestimates the frequency of high events in spring. The ten most biased days (considering regionally-averaged MDA8 O_3 values in each of the ten EPA regions) tend to be around 10°C warmer than average days. Our model does not include daily variations in U.S. anthropogenic emissions associated with higher electricity demand on hotter days (e.g., Abel et al., 2017), but we still find that the influence of U.S. anthropogenic emissions on regionally averaged MDA8 O_3 is up to 30 ppb higher on the ten most biased days as compared to average days. The model does include daily variability in temperature-sensitive biogenic emissions and simulates higher than average O_3 from BVOCs (up to 15 ppb) and soil NO_x (up to 10 ppb) on the ten most biased days. We conclude that regional production of O_3 is driving the pervasive high positive model bias in summer, as opposed to transported background.~~

On the ten days with the highest observed MDA8 O_3 values (O_3_top10obs) in each season, the model indicates that U.S. anthropogenic and biogenic VOC emissions are the most important drivers relative to average days, over most regions ~~(Tables 3, 4).~~(Table 2, Table 3). O_3_top10obs_MAM and O_3_top10obs_SON days (i.e., the ten highest spring and fall MDA8 O_3 days) are up to 9°C warmer than average, but O_3_top10obs_JJA days (i.e., the ten highest summer MDA8 O_3 days) are only 1-2 °C warmer than average~~.~~ (Table 2). U.S. anthropogenic emissions enhance O_3_top10obs_JJA days by 5-11 ppb above average in the eastern U.S. regions, but by less than 2 ppb over the three western regions. Over these westernmost regions, U.S. background O_3 is 4-12 ppb higher on O_3_top10obs_JJA days than on average~~.~~ (Table 2). Across the continental U.S.A., biogenic VOC emissions enhance O_3 by 1-7 ppb above average on O_3_top10obs_JJA days, while intercontinental pollution is either similar or up to 2 ppb higher on average days~~.~~ (Table 3). Analysis of our simulations thus indicates that the highest O_3 events are associated with regional O_3 production rather than transported background. ~~We emphasize, however, that our model is likely missing springtime events associated with stratospheric intrusions and Asian transport (Figure 3, Figure 7; Fiore et al., 2014; Zhang et al., 2011; 2014).~~

From 2004-2006 to 2010-2012, MDA8 O_3 concentrations on O_3_top10obs_JJA days vary from year-to-year, but show little overall trend ~~(decrease of~~, decreasing by 3 ppb in both the observations and the model averaged over all regions~~)~~ (Figure 6, ~~Table 5~~Table 4). With our sensitivity simulations, we interpret this lack of an overall trend as a balance between rising U.S. background O_3 (by 2 ppb for O_3_USB from 2004-2006 to 2010-2012 averaged over all

regions) and declining U.S. anthropogenic emissions (by 6 ppb for O_3_USA from 2004-2006 to 2010-2012 averaged over all regions). The declining influence of U.S. anthropogenic emissions on O_3_top10obs_JJA days is consistent with earlier work showing high-O_3 concentrations decreasing in response to regional precursor emissions controls since the late 1990s (e.g. Cooper et al., 2012, 2014b; Frost et al., 2006; Simon et al., 2016).

In contrast to previous work, including with the GEOS-Chem model (e.g. Fiore et al., 2014 and references therein), we find that U.S. background O_3 tends to be higher in summer than in spring in most regions. This likely reflects differences in the isoprene chemistry, specifically the isoprene nitrates, between our version of GEOS-Chem (Mao et al., 2013) and older versions that treat isoprene nitrates as greater sinks for NO_x and thereby, suppress O_3 production. ~~We find here that soil NO_x and isoprene can lead to high U.S. background O_3 in summer, though their relative importance is likely exaggerated at the coarse resolution we use here (e.g., Yu et al., 2016). Nevertheless, the model shows substantial variability in simulated U.S. background O_3 concentrations from year-to-year, on the order of 10-20 ppb between 2004-2012 in summer (Figure 11). The importance of temperature-sensitive sources like biogenic VOC and NO_x emissions to background O_3 imply that in a warmer climate, these background influences on O_3 will play an even more important role in driving up O_3~~The coarse resolution of our model will excessively mix isoprene and soil NO_x sources (e.g., Yu et al., 2016), and thus may exaggerate the relative importance of enhanced background O_3 resulting from soil NO_x and isoprene. Nevertheless, the model skill at capturing the observed year-to-year variability in the regionally averaged ten highest days lends some confidence to its attribution of this variability to natural sources (e.g. Figure 6). Future work with high-resolution models (e.g., at the regional scale, ideally with boundary conditions that include source attributions from a global model) is needed, along with observational evidence, to quantify the extent to which biogenic VOC and NO_x contribute to the highest observed O_3 levels in the warm season. The importance of temperature sensitive sources like biogenic VOC and NO_x emissions to background O_3 imply that in a warmer climate, these background influences on O_3 will play an even more important role in driving up O_3 levels.

Acknowledgments

We acknowledge insightful discussions with Gail Tonnesen and Pat Dolwick (U.S. EPA). We gratefully acknowledge support from NASA AQAST (NNX12AF15G) and NASA HAQAST (NNX16AQ20G). This project has been funded in part by the United States Environmental Protection Agency under assistance agreement RD83587801 to AMF. The contents of this document do not necessarily reflect the views and policies of the

Environmental Protection Agency, nor does the EPA endorse trade names or recommend the use of commercial products mentioned in this document. Upon publication, data behind the figures will be provided online at Columbia University Academic Commons.

References

Abatzoglou, J. T. and Williams, A. P.: Impact of anthropogenic climate change on wildfire across western US forests, Proc. Natl. Acad. Sci., 113(42), 11770–11775, doi:10.1073/pnas.1607171113, 2016.

~~Abel, D., Holloway, T., Kladar, R. M., Meier, P., Ahl, D., Harkey, M. and Patz, J.: Response of Power Plant Emissions to Ambient Temperature in the Eastern United States, Environ. Sci. Technol., 51(10), 5838–5846, doi:10.1021/acs.est.6b06201, 2017.~~

Auvray, M. and Bey, I.: Long-range transport to Europe: Seasonal variations and implications for the European ozone budget, J. Geophys. Res., 110(D11), D11303, doi:10.1029/2004JD005503, 2005.

Barkley, M. P., Palmer, P. I., Ganzeveld, L., Arneth, A., Hagberg, D., Karl, T., Guenther, A., Paulot, F., Wennberg, P. O., Mao, J., Kurosu, T. P., Chance, K., Müller, J. F., De Smedt, I., Van Roozendael, M., Chen, D., Wang, Y. and Yantosca, R. M.: Can a "state of the art" chemistry transport model simulate Amazonian tropospheric chemistry?, J. Geophys. Res. Atmos., 116(16), D16302, doi:10.1029/2011JD015893, 2011.

Baylon, P. M., Jaffe, D. A., Pierce, R. B. and Gustin, M. S.: Interannual Variability in Baseline Ozone and Its Relationship to Surface Ozone in the Western U.S., Environ. Sci. Technol., 50(6), 2994–3001, doi:10.1021/acs.est.6b00219, 2016.

Bond, T. C., Bhardwaj, E., Dong, R., Jogani, R., Jung, S., Roden, C., Streets, D. G. and Trautmann, N. M.: Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850-2000, Global Biogeochem. Cycles, 21(2), 1850–2000, doi:10.1029/2006GB002840, 2007.

Bosilovich, M. G.: Regional climate and variability of NASA MERRA and recent reanalyses: U.S. summertime precipitation and temperature, J. Appl. Meteorol. Climatol., 52(8), 1939–1951, doi:10.1175/JAMC-D-12-0291.1, 2013.

Bouwman, A. F., Lee, D. S., Asman, W. A. H., Dentener, F. J., Van Der Hoek, K. W. and Olivier, J. G. J.: A global high-resolution emission inventory for ammonia, Global Biogeochem. Cycles, 11(4), 561–587, doi:10.1029/97GB02266, 1997.

Clifton, O. E., Fiore, A. M., Munger, J. W., Malyshev, S., Horowitz, L. W., Shevliakova, E., Paulot, F., Murray, L. T. and Griffin, K. L.: Interannual variability in ozone removal by a temperate deciduous forest, Geophys. Res. Lett., 44(1), 542–552, doi:10.1002/2016GL070923, 2017.

Cooper, O. R., Gao, R.-S., Tarasick, D., Leblanc, T. and Sweeney, C.: Long-term ozone trends at rural ozone monitoring sites across the United States, 1990-2010, J. Geophys. Res. Atmos., 117(22), n/a-n/a, doi:http://dx.doi.org/10.1029/2012JD018261, 2012.

Cooper, O. R., Parrish, D. D., Ziemke, J., Balashov, N. V., Cupeiro, M., Galbally, I. E., Gilge, S., Horowitz, L., Jensen, N. R., Lamarque, J.-F., Naik, V., Oltmans, S. J., Schwab, J., Shindell, D. T., Thompson, A. M., Thouret, V., Wang, Y. and Zbinden, R. M.: Global distribution and trends of tropospheric ozone: An observation-based review, Elem. Sci. Anthr., 2, 29, doi:10.12952/journal.elementa.000029, 2014a.

Cooper, O. R., Parrish, D. D., Ziemke, J., Balashov, N. V., Cupeiro, M., Galbally, I. E., Gilge, S., Horowitz, L., Jensen, N. R., Lamarque, J.-F., Naik, V., Oltmans, S. J., Schwab, J., Shindell, D. T., Thompson, A. M., Thouret, V., Wang, Y. and Zbinden, R. M.: Global distribution and trends of tropospheric ozone: An observation-based review, Elem. Sci. Anthr., 2(0), 29, doi:10.12952/journal.elementa.000029, 2014b.

~~Van~~van Donkelaar, A., Martin, R. V, Leaitch, W. R., Macdonald, A. M., Walker, T. W., Streets, D. G., Zhang, Q., Dunlea, E. J., Jimenez, J. L., Dibb, J. E., Huey, L. G., Weber, R. and Andreae, M. O.: Analysis of aircraft and satellite measurements from the Intercontinental Chemical Transport Experiment (INTEX-B) to quantify long-range transport of East Asian sulfur to Canada, Atmos. Chem. Phys. Atmos. Chem. Phys., 8, 2999–3014 [online] Available from: www.atmos-chem-phys.net/8/2999/2008/ (Accessed 23 August 2016), 2008.

Fan, Y. and van den Dool, H.: A global monthly land surface air temperature analysis for 1948-present, J. Geophys. Res., 113(D1), 18, doi:10.1029/2007jd008470, 2008.

Fiore, A. M., Jacob, D. J., Field, B. D., Streets, D. G., Fernandes, S. D. and Jang, C.: Linking ozone pollution and climate change: The case for controlling methane, Geophys. Res. Lett., 29(19), 1919, doi:10.1029/2002GL015601, 2002.

Fiore, A. M., 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, J. Geophys. Res., 108(D24), 4787, doi:10.1029/2003jd003855, doi:10.1029/2003jd003855, 2003.

Fiore, A. M., Oberman, J. T., Lin, M., 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, Atmos. Environ., 96, 284–300, doi:10.1016/j.atmosenv.2014.07.045, 2014.

Fiore, A. M., Vaishali, N. and Leibensperger, E. M.: Air Quality and Climate Connections, J. Air Waste Manage. Assoc., 2015.
Free, M., Sun, B. and Yoo, H. L.: Comparison between total cloud cover in four reanalysis products and cloud measured by visual observations at U.S. weather stations, J. Clim., 29(6), 2015–2021, doi:10.1175/JCLI-D-15-0637.1, 2016.
Frost, G. J., McKeen, S. A., Trainer, M., Ryerson, T. B., Neuman, J. A., Roberts, J. M., Swanson, A., Holloway, J. S., Sueper, D. T., Fortin, T., Parrish, D. D., Fehsenfeld, F. C., Flocke, F., Peckham, S. E., Grell, G. A., Kowal, D., Cartwright, J., Auerbach, N. and Habermann, T.: Effects of changing power plant NOx emissions on ozone in the eastern United States: Proof of concept, J. Geophys. Res., 111(D12), D12306, doi:10.1029/2005JD006354, 2006.
Giglio, L., Randerson, J. T., Van Der Werf, G. R., Kasibhatla, P. S., Collatz, G. J., Morton, D. C. and Defries, R. S.: Assessing variability and long-term trends in burned area by merging multiple satellite fire products, Biogeosciences, 7(3), 1171–1186, doi:10.5194/bg-7-1171-2010, 2010.
Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K. and Wang, X.: The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5(6), 1471–1492, doi:10.5194/gmd-5-1471-2012, 2012.
Hu, L., Millet, D. B., Baasandorj, M., Griffis, T. J., Travis, K. R., Tessum, C. W., Marshall, J. D., Reinhart, W. F., Mikoviny, T., Müller, M., Wisthaler, A., Graus, M., Warneke, C. and de Gouw, J.: Emissions of C_6 -C_8 aromatic compounds in the United States: Constraints from tall tower and aircraft measurements, J. Geophys. Res. Atmos., 120(2), 826–842, doi:10.1002/2014JD022627, 2015.
~~Hudman, R. C., Moore, N. E., Mebust, A. K., Martin, R. V, Russell, A. R., Valin, L. C. and Cohen, R. C.: Steps towards a mechanistic model of global soil nitric oxide emissions: Implementation and space based-constraints, Atmos. Chem. Phys., 12(16), 7779–7795, doi:10.5194/acp-12-7779-2012, 2012.~~
Jacob, D. J., Horowitz, L. W., Munger, J. W., Heikes, B. G., Dickerson, R. R., Artz, R. S. and Keene, W. C.: Seasonal transition from NO\~ x-to hydrocarbon-limited conditions for ozone production over the eastern United States in September, J. Geophys. Res., 100, 9315–9315, doi:10.1029/94JD03125, 1995.
Jaffe, D. A.: Relationship between surface and free tropospheric ozone in the Western U.S., Environ. Sci. Technol., 45(2), 432–8, doi:10.1021/es1028102, 2011.
Jaffe, D. A., Cooper, O. R., Fiore, A. M., Henderson, B. H., Gail, S., Russell, A. G., Henze, D. K., Langford, A. O., Lin, M. and Moore, T.: Scientific assessment of background ozone over the U.S.: implications for air quality management, ~~2017~~[online] Available from: file:///C:/Users/shizi/Downloads/BOSA_complete_v6.pdf, 2018.
Kuhns, H. and Green, M.: Big Bend Regional Aerosol and Visibility Observational (BRAVO) Study Emissions Inventory, Desert Res. …, 89119(702) [online] Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.462.8648&rep=rep1&type=pdf (Accessed 29 January 2018), 2003.
Lee, C., Martin, R. V., van Donkelaar, A., Lee, H., Dickerson, R. R., Hains, J. C., Krotkov, N., Richter, A., Vinnikov, K. and Schwab, J. J.: SO_2 emissions and lifetimes: Estimates from inverse modeling using in situ and global, space-based (SCIAMACHY and OMI) observations, J. Geophys. Res., 116(D6), D06304, doi:10.1029/2010JD014758, 2011.
Leibensperger, E. M., Mickley, L. J., Jacob, D. J., Chen, W.-T., Seinfeld, J. H., Nenes, A., Adams, P. J., Streets, D. G., Kumar, N. and Rind, D.: Climatic effects of 1950–2050 changes in US anthropogenic aerosols – Part 2: Climate response, Atmos. Chem. Phys., 12(7), 3349–3362, doi:10.5194/acp-12-3349-2012, 2012.
Lin, M., Fiore, A. M., Cooper, O. R., Horowitz, L. W., Langford, A. O., Levy, H., Johnson, B. J., Naik, V., Oltmans, S. J. and Senff, C. J.: Springtime high surface ozone events over the western United States: Quantifying the role of stratospheric intrusions, J. Geophys. Res. Atmos., 117(D21), doi:10.1029/2012JD018151, 2012.
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., Nat. Commun., 6(May), 7105, doi:10.1038/ncomms8105, 2015a.
Lin, M., Horowitz, L. W., Cooper, O. R., Tarasick, D., Conley, S., Iraci, L. T., Johnson, B. J., Leblanc, T., Petropavlovskikh, I. and Yates, E. L.: Revisiting the evidence of increasing springtime ozone mixing ratios in the free troposphere over western North America, Geophys. Res. Lett., 42(20), 8719–8728, doi:10.1002/2015GL065311, 2015b.
Lin, M., Horowitz, L. W., Payton, R., Fiore, A. M. and Tonnesen, G.: US surface ozone trends and extremes from 1980-2014: Quantifying the roles of rising Asian emissions, domestic controls, wildfires, and climate, Atmos. Chem. Phys. Discuss., 1–56, doi:10.5194/acp-2016-1093, 2016.
Mao, J., Horowitz, L. W., Naik, V., Fan, S., Liu, J. and Fiore, A. M.: Sensitivity of tropospheric oxidants to biomass burning emissions: implications for radiative forcing, Geophys. Res. Lett., 40(6), 1241–1246,

doi:10.1002/grl.50210, 2013.
National Research Council: Global Sources of Local Pollution: An Assessment of Long-Range Transport of Key Air Pollutants to and from the United States, The National Academies Press, Washington, DC., 2010.
Olivier, J. G. J., Van Aardenne, J. A., Dentener, F. J., Pagliari, V., Ganzeveld, L. N. and Peters, J. A. H. W.: Recent trends in global greenhouse gas emissions:regional trends 1970–2000 and spatial distributionof key sources in 2000, Environ. Sci., 2(2–3), 81–99, doi:10.1080/15693430500400345, 2005.
Reidmiller, D. R., Fiore, A. M., Jaffe, D. A., Bergmann, D., Cuvelier, C., Dentener, F. J., Duncan, B. N., Folberth, G. A., Gauss, M., Gong, S., Hess, P., Jonson, J. E., Keating, T., Lupu, A., Marmer, E., Park, R. J., Schultz, M. G., Shindell, D. T., Szopa, S., Vivanco, M. G., Wild, O. and Zuber, A.: The influence of foreign vs. North American emissions on surface ozone in the US, Atmos. Chem. Phys., 9(14), 5027–5042, doi:10.5194/acp-9-5027-2009, 2009.
Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M. and Woollen, J.: MERRA: NASA's modern-era retrospective analysis for research and applications, J. Clim., 24(14), 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.
Schnell, J. L. and Prather, M. J.: Co-occurrence of extremes in surface ozone, particulate matter, and temperature over eastern North America., Proc. Natl. Acad. Sci. U. S. A., 114(11), 2854–2859, doi:10.1073/pnas.1614453114, 2017.
Schnell, J. L., Holmes, C. D., Jangam, A. and Prather, M. J.: Skill in forecasting extreme ozone pollution episodes with a global atmospheric chemistry model, Atmos. Chem. Phys., 14(15), 7721–7739, doi:10.5194/acp-14-7721-2014, 2014.
Schultz, M. G.: REanalysis of the TROpospheric chemical composition over the past 40 years, Reports Earth Syst. Sci., 48 [online] Available from: http://pubman.mpdl.mpg.de/pubman/item/escidoc:994467/component/escidoc:994466/WEB_BzE_48.pdf (Accessed 29 January 2018), 2007.
Simon, H., Wells, B., Baker, K. R. and Hubbell, B.: Assessing temporal and spatial patterns of observed and predicted ozone in multiple urban areas, Environ. Health Perspect., 124(9), 1443–1452, doi:10.1289/EHP190, 2016.
Stauffer, R. M., Thompson, A. M., Oltmans, S. J. and Johnson, B. J.: Tropospheric ozonesonde profiles at long-term U.S. monitoring sites: 2. Links between Trinidad Head, CA, profile clusters and inland surface ozone measurements, J. Geophys. Res. Atmos., 122(2), 1261–1280, doi:10.1002/2016JD025254, 2017.
Stettler, M. E. J., Eastham, S. and Barrett, S. R. H.: Air quality and public health impacts of UK airports. Part I: Emissions, Atmos. Environ., 45, 5415–5424, doi:10.1016/j.atmosenv.2011.07.012, 2011.
Streets, D. G., Zhang, Q., Wang, L., He, K., Hao, J., Wu, Y., Tang, Y. and Carmichael, G. R.: Revisiting China's CO emissions after the Transport and Chemical Evolution over the Pacific (TRACE-P) mission: Synthesis of inventories, atmospheric modeling, and observations, J. Geophys. Res. Atmos., 111(14), D14306, doi:10.1029/2006JD007118, 2006.
Task Force on Hemispheric Transport of Air Pollution: HEMISPHERIC TRANSPORT OF AIR POLLUTION, Econ. Comm. Eur. [online] Available from: http://www.htap.org/publications/2010_report/2010_Final_Report/HTAP 2010 Part A 110407.pdf (Accessed 30 June 2017), 2010.
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?, Atmos. Chem. Phys., 16(21), 13561–13577, doi:10.5194/acp-16-13561-2016, 2016.
Travis, K. R., Jacob, D. J., Keller, C. A., Kuang, S., Lin, J., Newchurch, M. J. and Thompson, A. M.: Resolving ozone vertical gradients in air quality models, Atmos. Chem. Phys. Discuss., 1–18, doi:10.5194/acp-2017-596, 2017.
U.S. Environmental Protection Agency: Map of EPA Regions, [online] Available from: http://www.epa.gov/oust/regions/regmap.htm (Accessed 2 December 2015), 2012.
U.S. Environmental Protection Agency: AirData - Download Data. [online] Available from: http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html#Daily, 2014.
Wang, C., Corbett, J. J. and Firestone, J.: Improving spatial representation of global ship emissions inventories, Environ. Sci. Technol., 42(1), 193–199, doi:10.1021/es0700799, 2008.
Wang, H., Jacob, D. J., Le Sager, P., Streets, D. G., Park, R. J., Gilliland, A. B. and van Donkelaar, A.: Surface ozone background in the United States: Canadian and Mexican pollution influences, Atmos. Environ., 43(6), 1310–1319, doi:10.1016/j.atmosenv.2008.11.036, 2009.

Van Der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., Defries, R. S., Jin, Y. and Van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997-2009), Atmos. Chem. Phys., 10(23), 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.

Xiao, Y., Logan, J. A., Jacob, D. J., Hudman, R. C., Yantosca, R. and Blake, D. R.: Global budget of ethane and regional constraints on U.S. sources, J. Geophys. Res., 113(D21), D21306, doi:10.1029/2007JD009415, 2008.

Yang, J., Tian, H., Tao, B., Ren, W., Pan, S., Liu, Y. and Wang, Y.: A growing importance of large fires in conterminous United States during 1984-2012, J. Geophys. Res. G Biogeosciences, 120(12), 2625–2640, doi:10.1002/2015JG002965, 2015.

Yevich, R. and Logan, J. A.: An assessment of biofuel use and burning of agricultural waste in the developing world, Global Biogeochem. Cycles, 17(4), n/a-n/a, doi:10.1029/2002GB001952, 2003.

Young, P. J., Naik, V., Fiore, A. M., Gaudel, A., Guo, J., Lin, M. Y., Neu, J., Parrish, D. D., Rieder, H. E., Schnell, J. L., Tilmes, S., Wild, O., Zhang, L., Brandt, J., Delcloo, A., Doherty, R. M., Geels, C., Hegglin, M. I., Hu, L., Im, U., Kumar, R., Luhar, A., Murray, L. T., Plummer, D., Rodriguez, J., Saiz-Lopez, A., Schultz, M. G., Woodhouse, M., Zeng, G. and Ziemke, J.: Tropospheric Ozone Assessment Report (TOAR): Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends, Elem. Sci. Anthr., 0–84 [online] Available from: http://eprints.lancs.ac.uk/88836/1/TOAR_Model_Performance_07062017.pdf (Accessed 12 December 2017), 2017.

Yu, K., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Miller, C. C., Travis, K. R., Zhu, L., Yantosca, R. M., Sulprizio, M. P., Cohen, R. C., Dibb, J. E., Fried, A., Mikoviny, T., Ryerson, T. B., Wennberg, P. O. and Wisthaler, A.: Sensitivity to grid resolution in the ability of a chemical transport model to simulate observed oxidant chemistry under high-isoprene conditions, Atmos. Chem. Phys, 16, 4369–4378, doi:10.5194/acp-16-4369-2016, 2016.

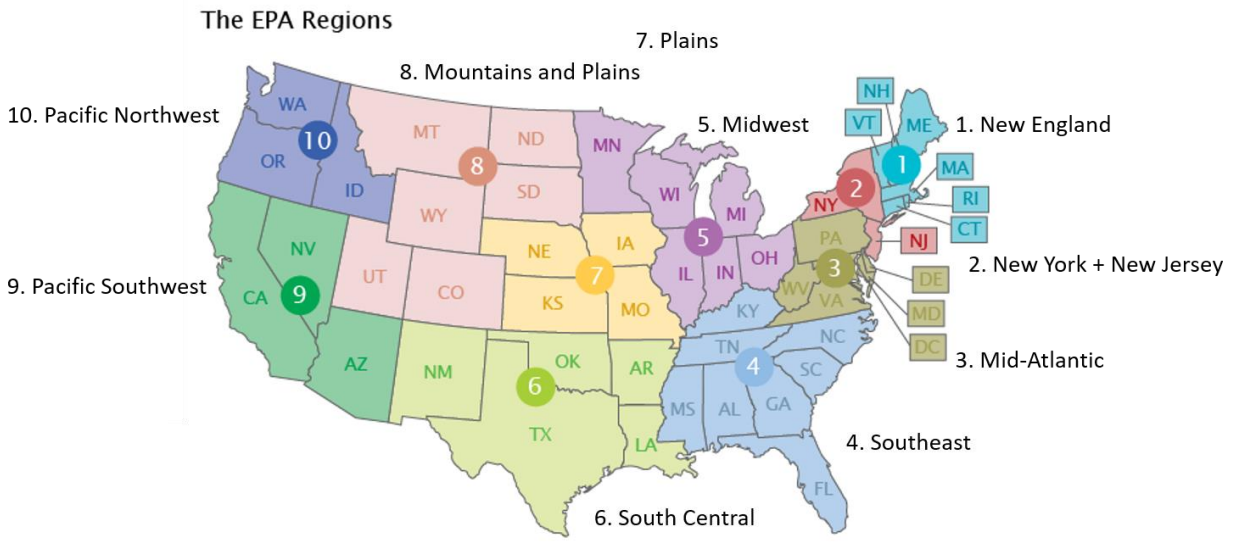
Zhang, L., Jacob, D. J., Downey, N. V., Wood, D. A., Blewitt, D., Carouge, C. C., van Donkelaar, A., Jones, D. B. A., Murray, L. T. and Wang, Y.: Improved estimate of the policy-relevant background ozone in the United States using the GEOS-Chem global model with 1/2° × 2/3° horizontal resolution over North America, Atmos. Environ., 45(37), 6769–6776, doi:10.1016/j.atmosenv.2011.07.054, 2011.

Zhang, L., Jacob, D. J., Yue, X., Downey, N. V., Wood, D. A. and Blewitt, D.: Sources contributing to background surface ozone in the US Intermountain West, Atmos. Chem. Phys., 14(11), 5295–5309, doi:10.5194/acp-14-5295-2014, 2014.

Figures

Table 1: ~~Approach for estimating~~Sensitivity simulations with the GEOS-Chem model and their application to estimate sources of ground-level O_3 ~~with the GEOS-Chem model~~.

Ozone Source	*Definition*	*Notation*
Base	Standard simulation	O_3_Base
Natural Background	Simulation with no global anthropogenic emissions + preindustrial CH_4 levels	O_3_NAT
North American Background	Simulation with no North American anthropogenic emissions	O_3_NAB
U.S. Background	Simulation with no U.S. anthropogenic emissions	O_3_USB
U.S. Anthropogenic Emissions	O_3_Base – O_3_USB	O_3_USA
Anthropogenic Emissions from Canada and Mexico	O_3_USB – O_3_NAB	O_3_CA+MX
Intercontinental Transport + Preindustrial CH_4 Levels	O_3_NAB – O_3_NAT	O_3_ICT+CH_4
North American Lightning NO_x	O_3_Base – simulation with the lightning NO_x source shut off	O_3_NALNO$_x$
Soil NO_x Emissions	O_3_Base – simulation with the soil NO_x emissions shut off	O_3_SNO$_x$
Terrestrial Biogenic VOC Emissions	O_3_Base – simulation with the terrestrial biogenic emissions shut off	O_3_BVOC
All Emissions except Terrestrial Biogenic VOCs	~~No~~Simulation with terrestrial biogenic VOC emissions shut off	O_3_noBVOC
Biomass Burning Emissions	O_3_Base – simulation with biomass burning emissions (NO_x, CO, VOCs, aerosols, and precursors from fires) shut off	O_3_BB



~~**Figure 1: Map of the states falling within each EPA region in the continental United States (adapted from *U.S. Environmental Protection Agency*, 2012).**~~

~~***Table 2*: The number of observational sites that fall within each EPA region for EPA AQS and CASTNet. (*) We include data from the Mount Bachelor Observatory in the Pacific Northwest region.**~~

~~**Region**~~	~~**EPA AQS**~~	~~**CASTNet**~~	~~***Total***~~
~~1. New England~~	~~82~~	~~7~~	~~*89*~~
~~2. New York + New Jersey (NY+NJ)~~	~~61~~	~~7~~	~~*68*~~
~~3. Mid-Atlantic~~	~~138~~	~~14~~	~~*152*~~
~~4. Southeast~~	~~309~~	~~24~~	~~*333*~~
~~5. Midwest~~	~~255~~	~~18~~	~~*273*~~
~~6. South Central~~	~~202~~	~~5~~	~~*207*~~
~~7. Plains~~	~~71~~	~~2~~	~~*73*~~
~~8. Mountains and Plains~~	~~153~~	~~12~~	~~*165*~~
~~9. Pacific Southwest~~	~~325~~	~~14~~	~~*339*~~
~~10. Pacific Northwest~~	~~48~~	~~6*~~	~~*54*~~
~~***Total***~~	~~***1644***~~	~~***109***~~	~~***1753***~~

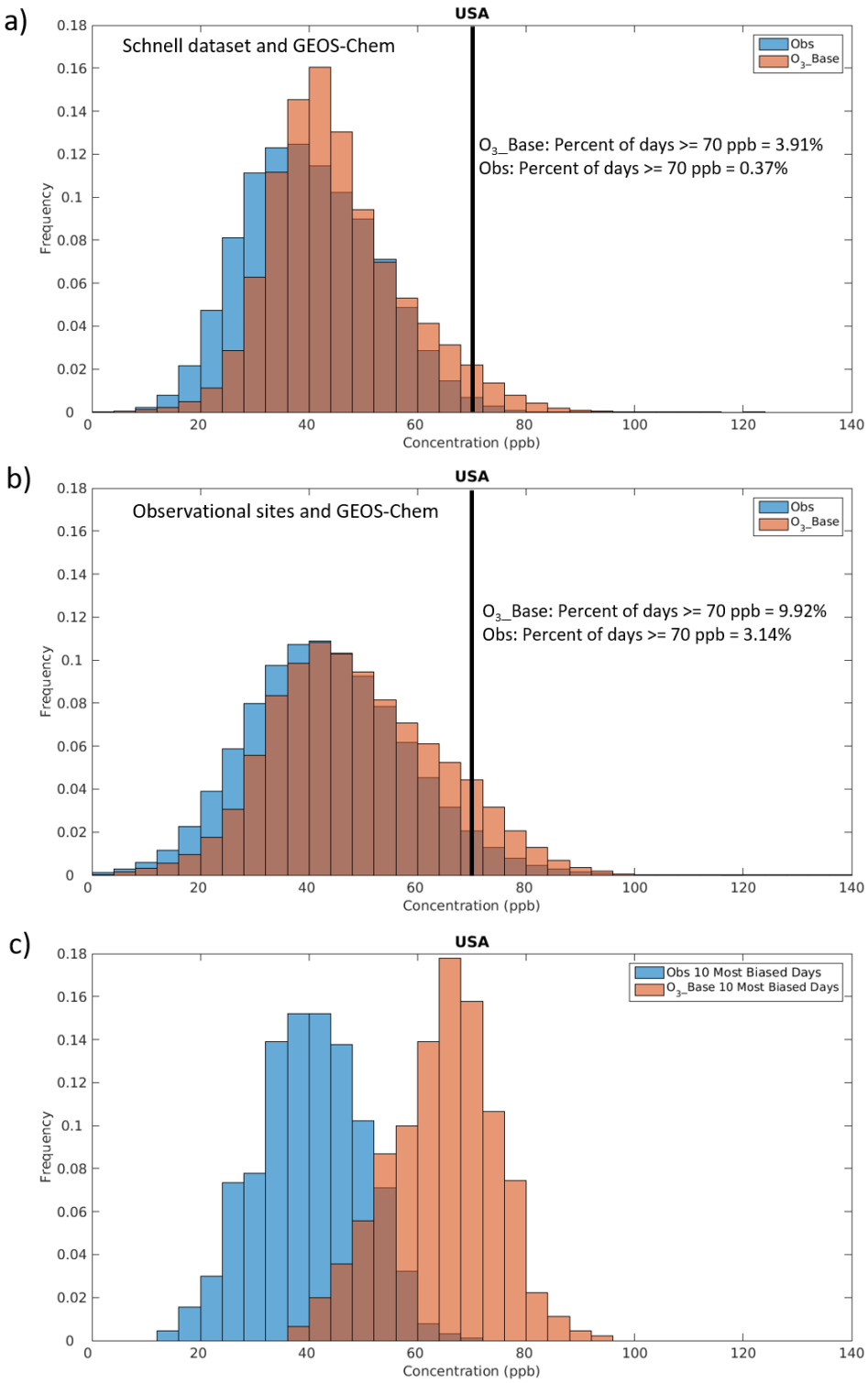
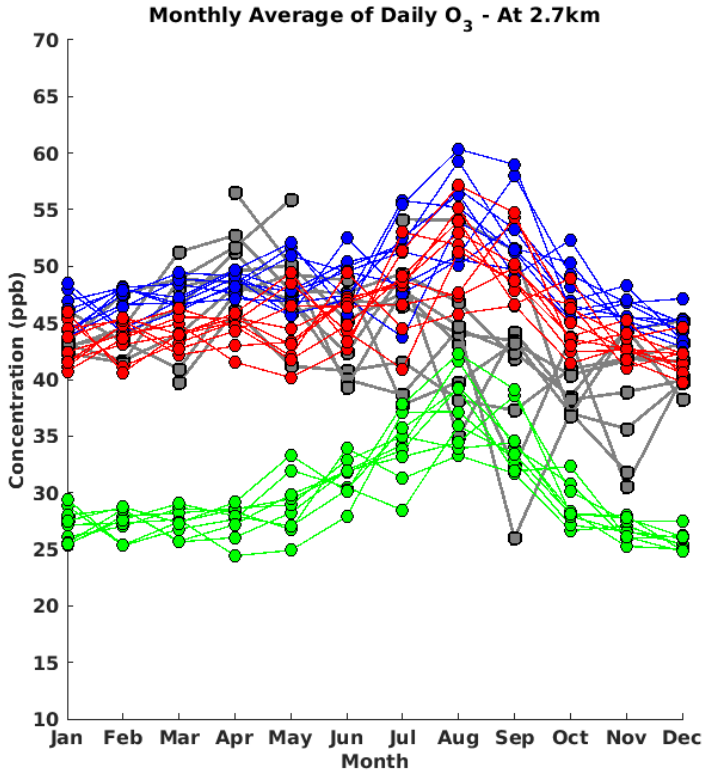


Figure 1~~: Frequency distribution of MDA8 O_3 values across all sites in the United States from Jan-Dec (365 or 366 days per year) from 2004-2012 in the (a) Schnell dataset (2014) interpolated to 2° by 2.5°, (b) at individual observational sites, and c) on the 10 most biased days. Concentrations for each day are obtained by averaging across all sites in a region. The model bias is defined as O_3_Base minus observed. The total number of points consists of 9 years x 10 days x 10 regions. The observations are in shown in blue and GEOS-Chem is in orange. The line drawn at 70 ppb in panels (a) and (b) denotes the 70 ppb NAAQS standard cut-off for O_3.~~: Frequency distribution of regionally averaged U.S. MDA8 O_3 values from 2004-

2012 in the (a) Schnell and Prather (2017) dataset interpolated to 2° by 2.5° and (b) at individual observational sites prior to averaging over each of the 10 EPA regions (total number of points is 9 years x 365 or 366 days x 10 regions) in the observations (blue) and the GEOS-Chem model (orange). c) As in panel (b) but selecting for the 10 most biased days in each region (total number of points is 9 years x 10 days x 10 regions). The line drawn at 70 ppb in panels (a) and (b) is the current O_3 NAAQS level.



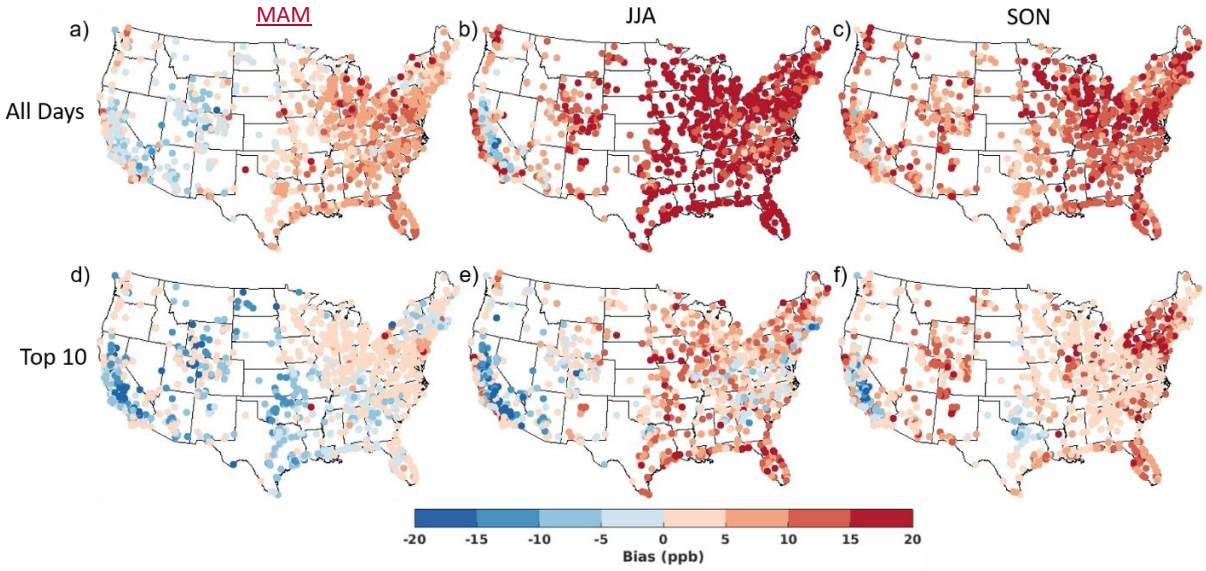
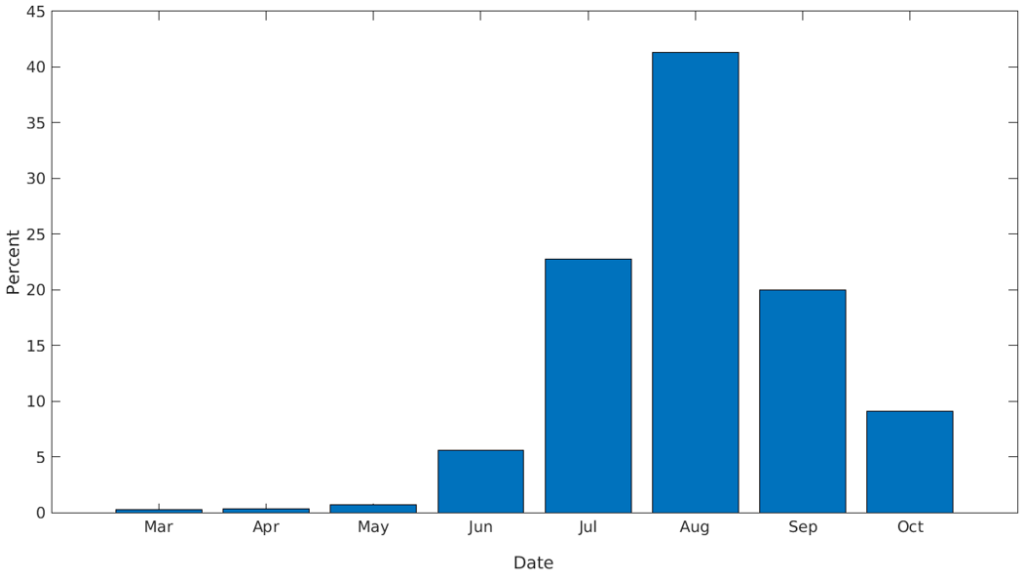


Figure 2

: Average MDA8 O_3 model bias (O_3_Base – observed) on all days in (a) ~~*JJA, (b)*~~ MAM, (b) JJA, and (c) SON versus on the (d) O_3_top10obs_MAM, (e) O_3_top10obs_JJA, and (f) O_3_top10obs_SON days at each observational site averaged across 2004-2012. ~~: Monthly average concentrations of daily O_3 at Mount Bachelor Observatory (Observations; grey), with~~

~~corresponding O_3_Base (blue), O_3_USB (red), and O_3_NAT (green) concentrations at ~2.7 km, the height of the Mount Bachelor Observatory. Individual lines of the same color show the spread from 2004-2012.~~

~~*Figure 4:*~~



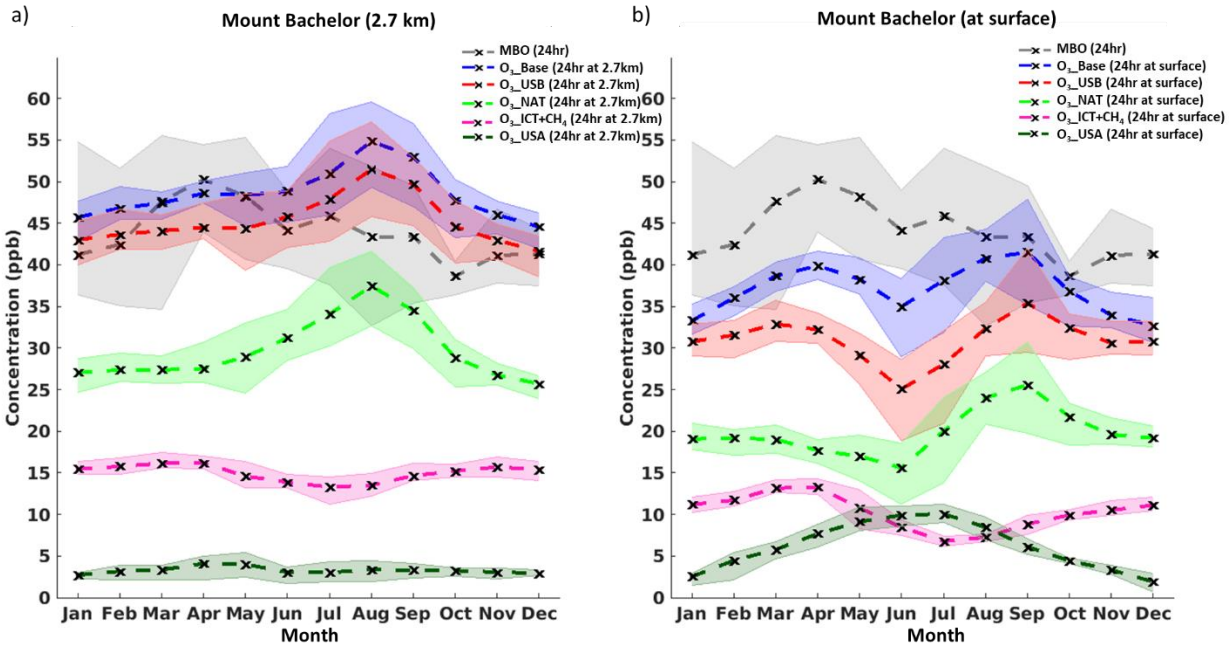


Figure 3: Monthly 2004-2012 average 24-hour O_3 concentrations at Mount Bachelor Observatory. Observations (grey) are the same in both panels. Simulations from the GEOS-Chem model are sampled in the grid cell containing Mount Bachelor at (a) 2.7 km (the height of the Mount Bachelor Observatory) and at (b) the surface: O_3_Base (blue), O_3_USB (red), O_3_NAT (light green), O_3_ICT+CH4 (pink), and O_3_USA (dark green). The shaded range spans the highest and lowest years. ~~: Percent of total top 10 most biased days from Jan-Dec (9 years x 10 days x 10 regions) that fell within each month in the United States. All the most biased days fell between Mar-Oct.~~

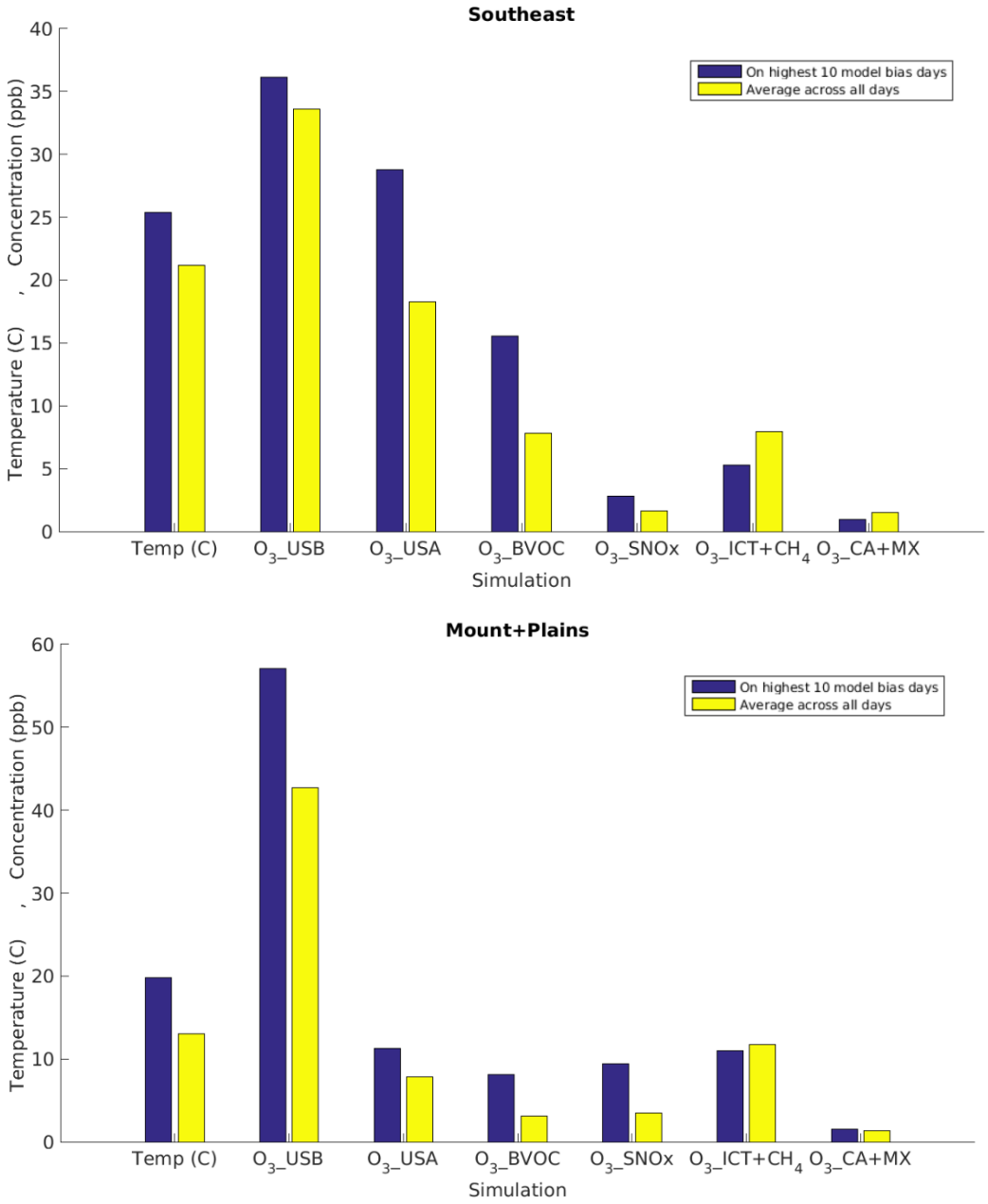
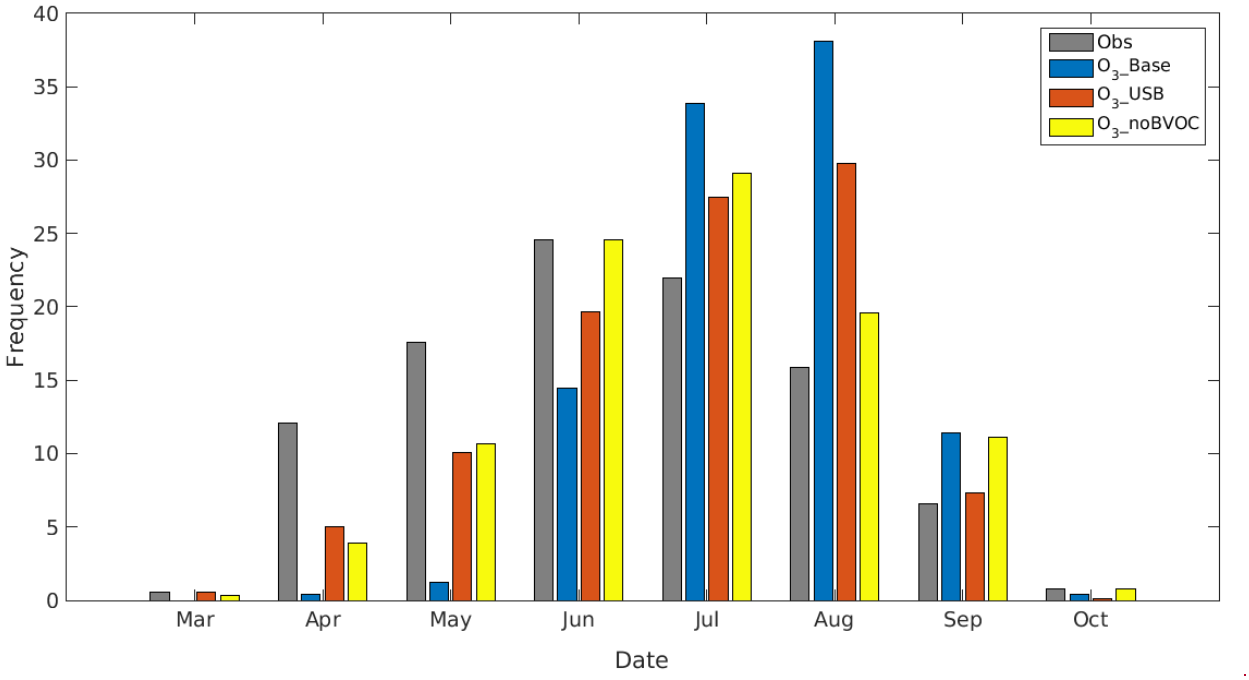
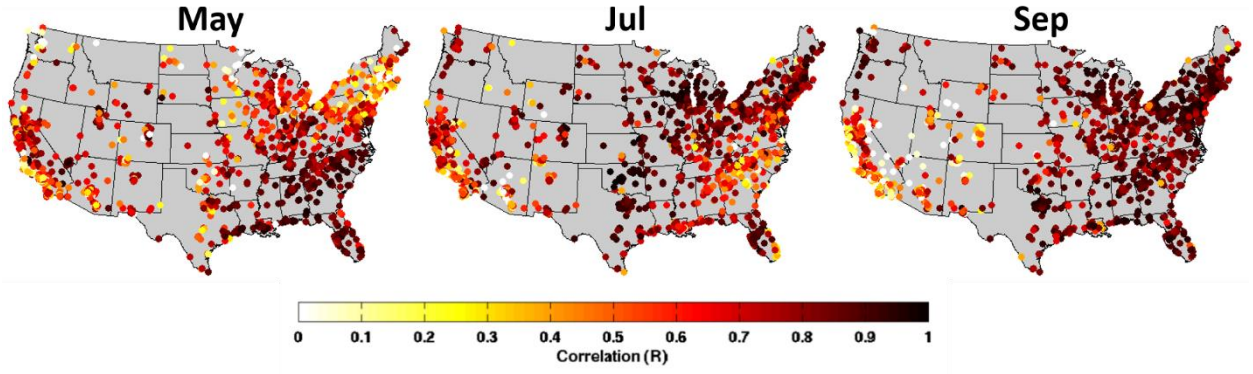


Figure 4~~: Average influence of each sensitivity simulation on MDA8 O_3~~: Multi-year (2004-2012) Mar-Oct average temperature and MDA8 O_3 source contributions estimated with the GEOS-Chem model in the (a) Southeast and (b) Mountain and Plains regions on the 10 most biased days ~~from Jan-Dec (red~~(blue) versus averaged across all days (~~blue). Red circles show the average model bias (O_3_Base – observations) on the 10 most biased days. Blue circles show the model bias averaged across all days. The circles do not vary between subplots.~~yellow). Note that ~~O_3_USB and O_3_USA~~the two regions are on ~~a~~ different ~~scale than the other plots~~scales.



~~**Figure 7: Percent of total top ten days (9 years x 10 days x 10 regions) from Jan-Dec (365 or 366 days) in the observations, O_3_Base, O_3_USB, and O_3_noBVOC that fell within each month for all sites across the U.S.A. All the top ten days for each simulation fell between Mar-Oct.**~~



~~**Figure 8: Correlation between 2004-2012 year-to-year monthly MDA8 O_3 averages for May, July, and September in the observation and in the model (O_3_Base).**~~

Table 2: Summary information for each region. The "Model Bias" column shows the model bias in each region on the (1) O_3_top10obs days in each season (average of 2004-2012), 2) across all days in each season (average of 2004-2012), and (3) the difference between these values, rounded to the nearest whole number. The other columns show the concentration for the observations, O_3_Base, and O_3_USA, and daily average temperature (in degrees C) on the (1) O_3_top10obs days in each season (average of 2004-2012), (2) across all days in each season (average of 2004-2012), and (3) the difference between these values.

	Metric	Model Bias			O_3_Base			Obs			O_3_USB			O_3_USA			Temperature (C)		
Region	*Season*	*MAM*	*JJA*	*SON*	*MAM*	*JJA*	*SON*	*MAM*	*JJA*	*SON*	*MAM*	*JJA*	*SON*	*MAM*	*JJA*	*SON*	*MAM*	*JJA*	*SON*
New England	Top 10 Days	-1	9	12	57	72	59	58	62	59	35	39	36	22	33	23	12	22	18
	Avg all days	2	14	10	46	56	41	44	43	31	35	35	32	12	21	9	7	19	11
	Difference	**-3**	**-4**	**2**	**11**	**16**	**18**	**13**	**20**	**28**	**0**	**4**	**3**	**10**	**11**	**14**	**5**	**2**	**6**
NY+NJ	Top 10 Days	4	14	15	63	80	65	59	67	65	36	42	38	27	38	27	16	24	20
	Avg all days	5	18	11	49	65	42	44	47	31	35	39	33	14	27	10	9	21	12
	Difference	**-1**	**-5**	**4**	**14**	**15**	**23**	**15**	**19**	**34**	**1**	**3**	**6**	**13**	**11**	**18**	**6**	**2**	**7**
Mid-Atlantic	Top 10 Days	3	13	15	65	81	69	63	68	69	34	40	37	31	40	32	18	25	21
	Avg all days	6	18	12	51	70	45	46	52	33	34	38	32	17	32	13	12	23	14
	Difference	**-3**	**-5**	**3**	**14**	**11**	**24**	**17**	**16**	**36**	**0**	**2**	**5**	**14**	**9**	**19**	**6**	**2**	**7**
Southeast	Top 10 Days	0	13	10	62	72	63	61	59	63	34	39	34	27	34	28	19	26	21
	Avg all days	6	19	12	55	65	51	48	46	39	34	37	32	21	28	19	17	26	18
	Difference	**-6**	**-6**	**-1**	**7**	**7**	**12**	**13**	**13**	**24**	**1**	**2**	**3**	**7**	**5**	**10**	**2**	**1**	**3**
Midwest	Top 10 Days	4	14	17	63	77	70	59	63	70	36	44	42	27	33	27	17	24	21
	Avg all days	6	19	11	49	68	43	44	48	32	34	42	33	15	26	10	10	22	12
	Difference	**-1**	**-5**	**6**	**14**	**10**	**26**	**15**	**15**	**37**	**1**	**2**	**9**	**12**	**8**	**17**	**7**	**1**	**9**
South Central	Top 10 Days	0	13	9	60	75	67	60	62	67	39	45	40	21	30	26	20	27	23
	Avg all days	5	17	10	52	62	51	47	46	41	36	41	35	16	21	16	18	27	19
	Difference	**-5**	**-4**	**-2**	**8**	**12**	**15**	**14**	**16**	**26**	**3**	**4**	**5**	**6**	**9**	**10**	**2**	**1**	**4**
Plains	Top 10 Days	0	13	13	58	74	67	58	61	67	37	47	42	21	28	25	17	26	22
	Avg all days	5	18	10	50	67	45	44	49	35	34	44	34	15	23	11	13	25	13
	Difference	**-6**	**-5**	**3**	**8**	**8**	**23**	**14**	**13**	**33**	**2**	**3**	**9**	**6**	**5**	**14**	**4**	**1**	**9**
Mountains + Plains	Top 10 Days	-1	8	13	56	69	64	57	60	64	45	57	54	11	12	10	12	22	18
	Avg all days	0	11	9	50	64	48	50	53	39	41	53	41	10	11	7	7	20	9
	Difference	**-1**	**-2**	**4**	**6**	**5**	**16**	**7**	**7**	**25**	**5**	**4**	**12**	**1**	**0**	**3**	**5**	**2**	**9**
Pacific SW	Top 10 Days	-3	3	6	57	64	63	60	62	63	41	47	48	16	18	15	18	25	24
	Avg all days	0	4	8	49	57	49	49	53	42	37	41	39	12	16	10	14	23	17
	Difference	**-3**	**-1**	**-2**	**8**	**7**	**14**	**10**	**9**	**21**	**4**	**6**	**8**	**4**	**2**	**5**	**5**	**2**	**7**
Pacific NW	Top 10 Days	-1	6	11	48	59	51	49	52	51	39	49	44	9	10	7	12	22	17
	Avg all days	2	8	10	43	46	40	41	38	30	35	36	36	8	10	4	8	17	10
	Difference	**-3**	**-2**	**0**	**5**	**13**	**11**	**9**	**14**	**21**	**4**	**12**	**9**	**1**	**0**	**3**	**4**	**4**	**7**

Table 3: Summary information for each region. Each column shows the concentration for each background O_3 source influence on the (1) O_3_top10obs days in each season (average of 2004-2012), (2) across all days in each season (average of 2004-2012), and (3) the difference between these values, rounded to the nearest whole number.

	Metric	O_3_USB			O_3_BVOC			O_3_SNO_x			O_3_$NALNO_x$			O_3_ICT+CH_4			O_3_CA+MX		
Region	**Season**	**MAM**	**JJA**	**SON**	**MAM**	**JJA**	**SON**	**MAM**	**JJA**	**SON**	**MAM**	**JJA**	**SON**	**MAM**	**JJA**	**SON**	**MAM**	**JJA**	**SON**
New England	Top 10 Days	35	39	36	6	17	13	1	3	2	1	2	1	8	3	5	7	7	5
	Avg all days	35	35	32	2	10	6	1	3	2	1	2	2	10	4	7	6	6	4
	Difference	**0**	**4**	**3**	**4**	**7**	**8**	**0**	**0**	**1**	**0**	**0**	**0**	**-2**	**-1**	**-3**	**1**	**1**	**2**
NY+NJ	Top 10 Days	36	42	38	9	20	17	1	4	3	1	2	2	7	2	4	6	6	5
	Avg all days	35	39	33	3	14	7	1	3	2	1	2	2	10	4	7	5	6	4
	Difference	**1**	**3**	**6**	**6**	**6**	**9**	**0**	**0**	**1**	**0**	**0**	**0**	**-2**	**-1**	**-3**	**1**	**0**	**2**
Mid-Atlantic	Top 10 Days	34	40	37	10	20	18	1	4	3	1	3	2	7	3	5	4	3	4
	Avg all days	34	38	32	5	16	9	1	3	2	1	3	2	9	4	7	4	4	3
	Difference	**0**	**2**	**5**	**5**	**4**	**9**	**0**	**1**	**1**	**0**	**0**	**0**	**-2**	**-1**	**-2**	**0**	**0**	**1**
Southeast	Top 10 Days	34	39	34	7	16	14	2	4	2	1	3	2	8	4	6	2	2	2
	Avg all days	34	37	32	5	14	9	1	3	2	2	4	2	9	5	7	2	1	2
	Difference	**1**	**2**	**3**	**2**	**2**	**4**	**0**	**1**	**1**	**-1**	**-1**	**0**	**0**	**-1**	**-1**	**0**	**1**	**0**
Midwest	Top 10 Days	36	44	42	8	16	16	2	6	5	1	2	2	6	1	4	3	3	3
	Avg all days	34	42	33	3	13	8	1	6	2	1	2	2	9	2	6	4	3	2
	Difference	**1**	**2**	**9**	**4**	**3**	**8**	**1**	**0**	**2**	**0**	**0**	**0**	**-3**	**-1**	**-3**	**0**	**0**	**1**
South Central	Top 10 Days	39	45	40	6	17	14	3	5	4	2	4	2	8	5	6	3	2	2
	Avg all days	36	41	35	4	12	8	2	4	2	2	6	2	9	7	8	3	2	2
	Difference	**3**	**4**	**5**	**2**	**5**	**6**	**1**	**1**	**2**	**0**	**-2**	**0**	**-1**	**-2**	**-2**	**0**	**0**	**0**
Plains	Top 10 Days	37	47	42	5	16	14	3	8	6	1	3	2	7	2	4	3	1	1
	Avg all days	34	44	34	3	13	7	2	8	3	1	3	2	8	3	7	3	2	2
	Difference	**2**	**3**	**9**	**2**	**3**	**7**	**1**	**0**	**3**	**0**	**0**	**0**	**-1**	**-1**	**-2**	**0**	**0**	**0**
Mountains + Plains	Top 10 Days	45	57	54	1	8	7	3	9	7	3	5	5	12	11	12	2	1	1
	Avg all days	41	53	41	1	7	4	2	8	3	2	5	4	12	11	12	2	2	1
	Difference	**5**	**4**	**12**	**1**	**1**	**3**	**1**	**1**	**4**	**1**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**0**
Pacific SW	Top 10 Days	41	47	48	3	9	9	3	5	5	2	4	4	11	8	10	2	2	2
	Avg all days	37	41	39	1	7	5	1	4	3	2	4	3	12	8	11	1	2	2
	Difference	**4**	**6**	**8**	**2**	**2**	**4**	**1**	**1**	**2**	**0**	**0**	**1**	**-1**	**0**	**-1**	**0**	**0**	**0**
Pacific NW	Top 10 Days	39	49	44	0	9	7	2	7	5	1	3	3	12	9	10	3	4	2
	Avg all days	35	36	36	-1	4	3	1	4	2	1	2	3	13	9	10	2	3	1
	Difference	**4**	**12**	**9**	**1**	**5**	**4**	**1**	**3**	**3**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**1**

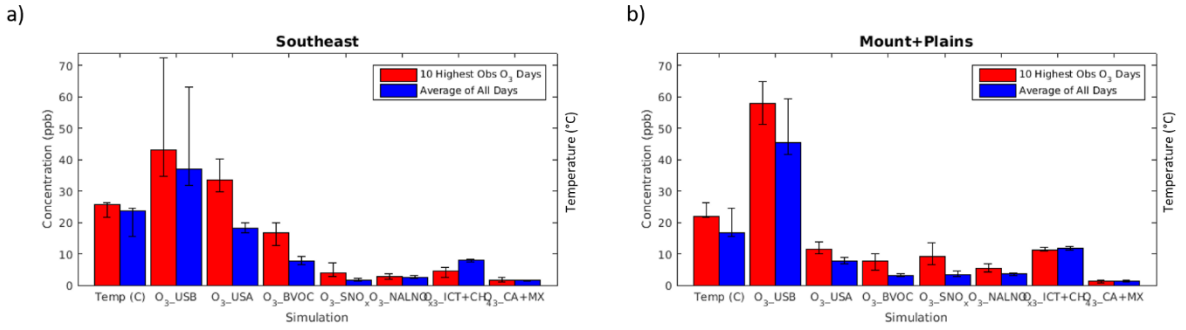


Figure 5: Average 2004-2012 influence of each sensitivity simulation to O_3_Base in the (a) Southeast and (b) Mountains and Plains regions on MDA8 O_3_top10obs_JJA days (red) versus averaged across all days (blue). Error bars show the concentration on the lowest versus highest year for each sensitivity simulation in each region.

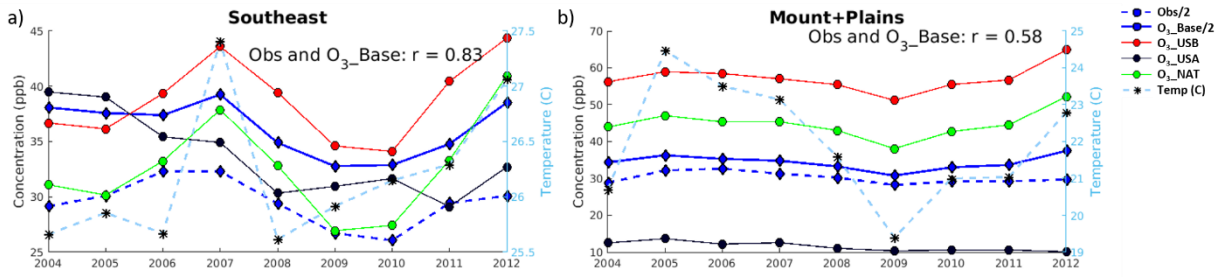
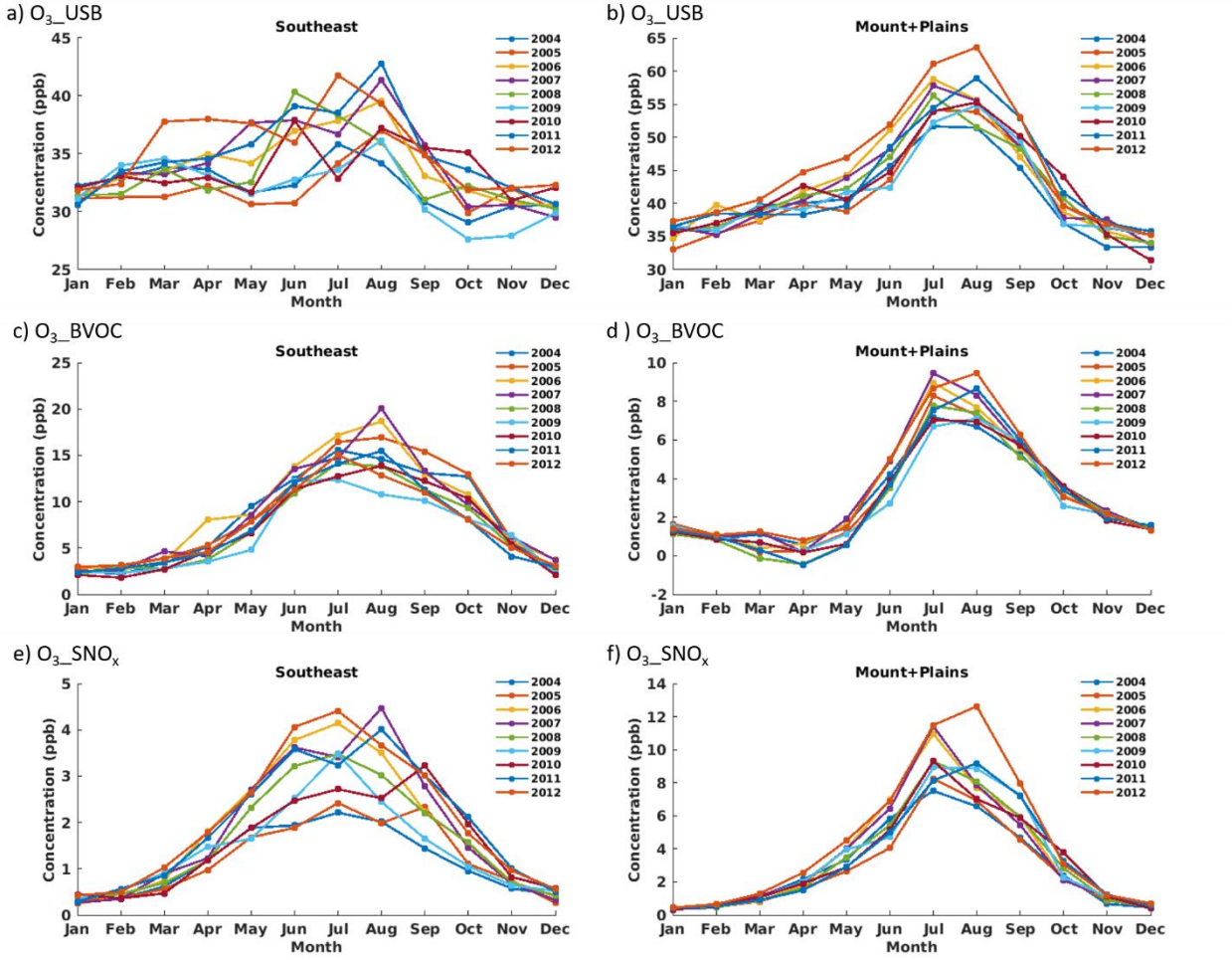


Figure 6: Average yearly MDA8 O_3_top10obs_JJA concentrations for observations (divided by 2 to fit on the same axes; blue dashed line), O_3_Base (divided by 2; blue solid line), O_3_USB (red), O_3_USA (black), O_3_NAT (green) MDA8, and daily average temperature (in degrees C; light blue) in the (a) Southeast and (b) Mountains and Plains regions.

Table 4: Change in MDA8 O_3 concentrations from 2004-2006 to 2010-2012 on O_3_top10obs_JJA days in the observations, O_3_Base, O_3_USB, ~~*and*~~ O_3_USA, and temperature.

	Obs	O_3_Base	O_3_USB	O_3_USA	Temperature (C)
New England	-6	-4	6	-10	2
NY+NJ	-2	-4	3	-7	1
Mid-Atlantic	0	-3	4	-7	1
Southeast	-4	-5	2	-7	1
Midwest	-2	-4	2	-6	0
South Central	-6	-2	5	-7	1
Plains	-1	-2	4	-5	1
Mountains + Plains	-4	-1	1	-2	-1
Pacific SW	-3	-4	0	-4	-1
Pacific NW	-7	-5	-4	-1	-1
Average	***-3***	***-3***	***2***	***-6***	***0***



900

Figure 7: Monthly average MDA8 O_3_USB (a, b), O_3_BVOC (c, d), and O_3_SNO_x (e, f) concentrations in the Southeast (a, c, e) and Mountains and Plains (b, d, f) regions.

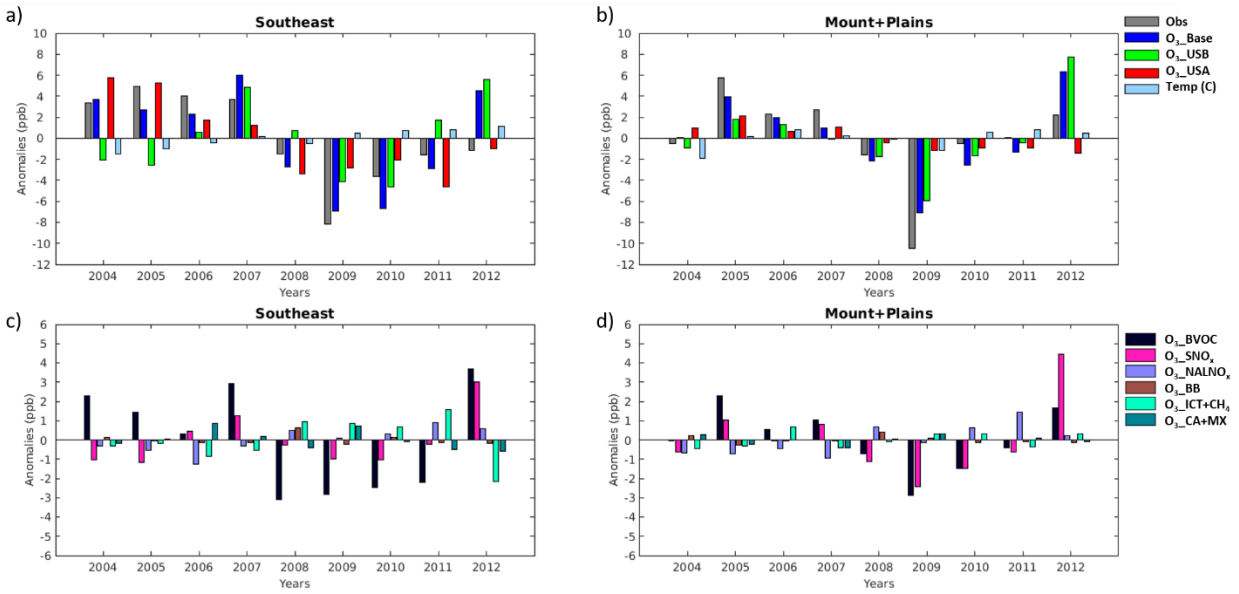
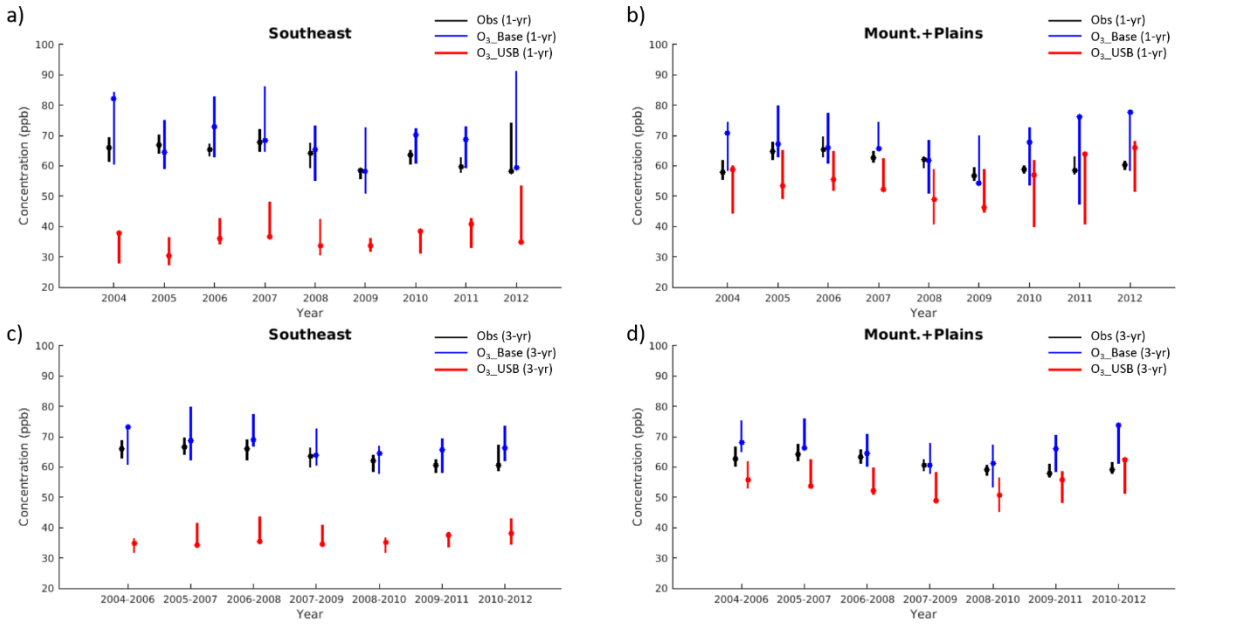


Figure 8: Anomaly on the MDA8 O_3_top10obs_JJA days relative to the 2004-2012 average in the Southeast (a, c) and in the Mountains and Plains (b, d) regions. Panels (a) and (b) show the observations, O_3_Base, O_3_USB, O_3_USA, and temperature (in degrees C). Panels (c) and (d) show O_3_BVOC, O_3_SNO_x, O_3_$NALNO_x$, O_3_BB, O_3_ICT+CH_4, and O_3_CA+MX.



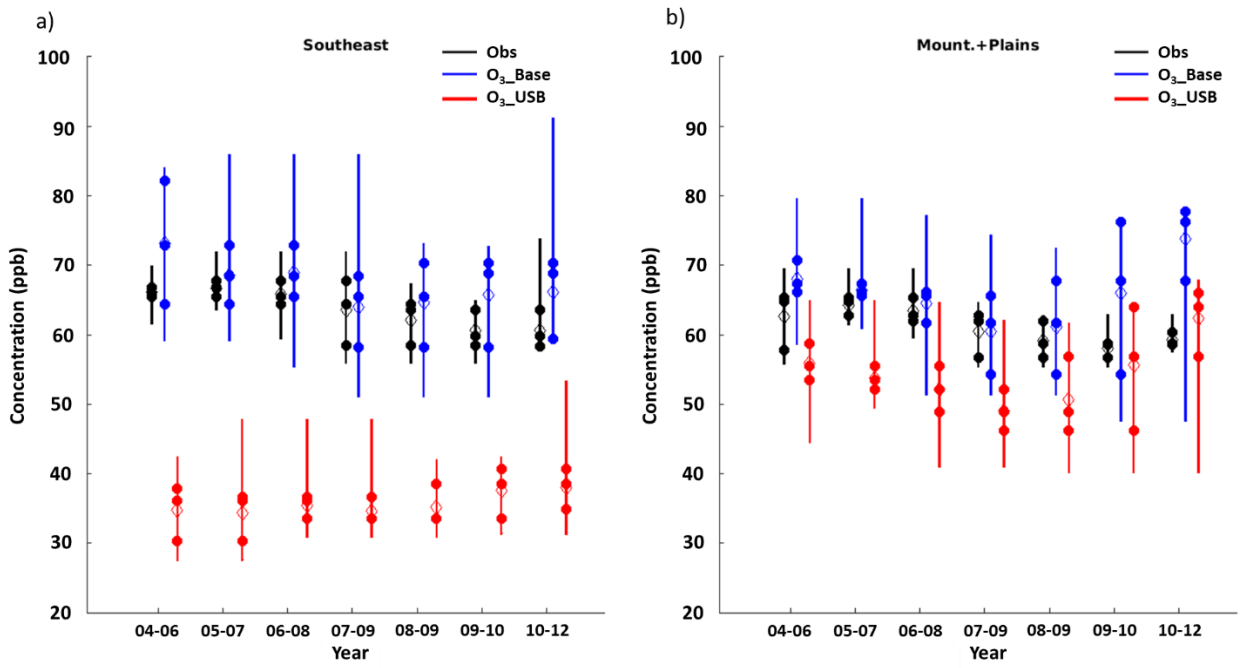


Figure 9~~**: Range**~~**: The three 4th highest days in** ~~**magnitude of the MDA8 O3_top10obs for**~~ **each year** ~~**shown as vertical lines**~~**(solid dots) that went into the calculation of the three-year average of the 4th highest MDA8 O3 day (hollow diamond). Error bars show the range between the highest and lowest O3_top10obs days across each 3-year span (i.e, across 30 total points) occurring between March and October in the (a) Southeast and (b) Mountains and Plains regions in the observations (black), and the O3_Base (blue)**~~**,**~~ **and O3_USB (red)** ~~**in the (a, c) Southeast and (b, d) Mountains and Plains regions. (a, b) show the range on of O3_top10obs days during each year between 2004-2012. (c, d) show**~~**simulations sampled on the same days as the top 10 observed values.**

Table 5~~**: Summary information for each region. The first row next to each region reports**~~ **the range across 2004-2012 of the** ~~**O3_top10obs days after averaging over three consecutive**~~**4th highest values from each of the 9 individual years**~~**. *The solid dots show the 4th highest MDA8 O3 day for each simulation (a, b) and the annual 4th highest* MDA8**~~ **for the observations, O3_Base, and O3_USB. The second row reports the range across 2004-2012 of each of the 3-year averages of the 4th highest values (7 values) in each region for the observations, O3** ~~**day averaged over three consecutive years.**~~

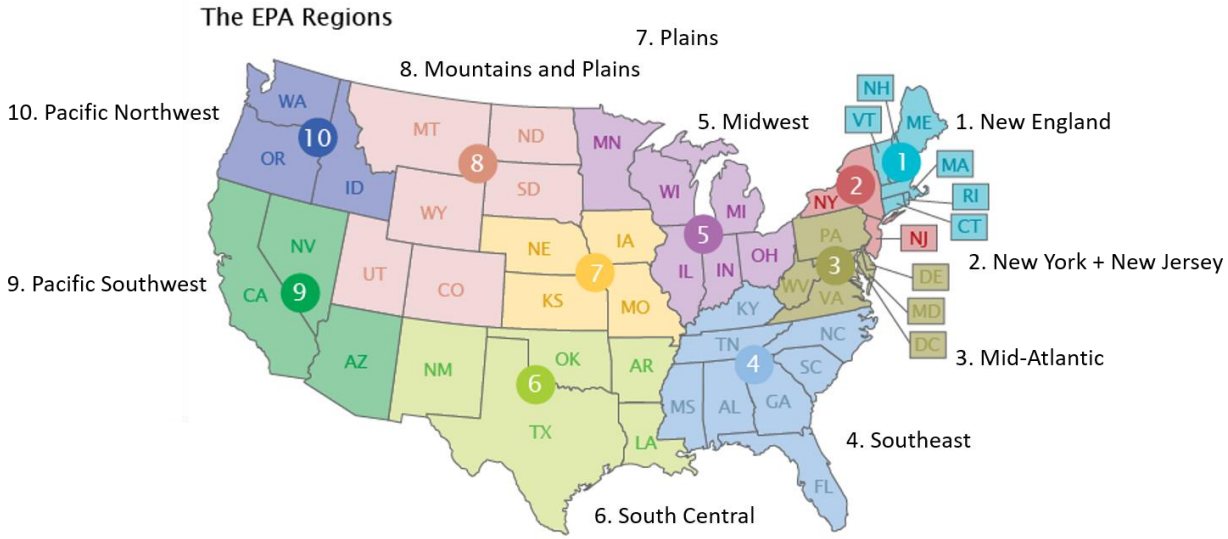
_Base, and O_3_USB.

Region	*Range*	Obs	O_3_Base	O_3_USB
New England	4th highest day	15	16	10
	3-year average 4th highest day	9	10	3
	Difference	**-6**	**-6**	**-7**
NY+NJ	4th highest day	11	10	12
	3-year average 4th highest day	6	2	6
	Difference	**-5**	**-8**	**-6**
Mid-Atlantic	4th highest day	13	36	25
	3-year average 4th highest day	7	21	10
	Difference	**-6**	**-15**	**-15**
Southeast	4th highest day	9	24	10
	3-year average 4th highest day	6	9	4
	Difference	**-3**	**-15**	**-7**
Midwest	4th highest day	13	22	24
	3-year average 4th highest day	8	11	10
	Difference	**-6**	**-11**	**-14**
South Central	4th highest day	11	26	22
	3-year average 4th highest day	8	13	13
	Difference	**-3**	**-13**	**-9**
Plains	4th highest day	14	32	24
	3-year average 4th highest day	9	18	11
	Difference	**-5**	**-15**	**-13**
Mountains + Plains	4th highest day	9	23	20
	3-year average 4th highest day	6	13	13
	Difference	**-2**	**-10**	**-7**
Pacific SW	4th highest day	5	23	20
	3-year average 4th highest day	3	5	5
	Difference	**-2**	**-18**	**-15**
Pacific NW	4th highest day	11	14	15
	3-year average 4th highest day	5	9	12
	Difference	**-5**	**-5**	**-3**

Supplemental Figures

Supplemental Table 1: The number of observational sites that fall within each EPA region for EPA AQS and CASTNet. (*) We include data from the Mount Bachelor Observatory in the Pacific Northwest region.

Region	EPA AQS	CASTNet	*Total*
1. New England	82	7	*89*
2. New York + New Jersey (NY+NJ)	61	7	*68*
3. Mid-Atlantic	138	14	*152*
4. Southeast	309	24	*333*
5. Midwest	255	18	*273*
6. South Central	202	5	*207*
7. Plains	71	2	*73*
8. Mountains and Plains	153	12	*165*
9. Pacific Southwest	325	14	*339*
10. Pacific Northwest	48	6*	*54*
Total	***1644***	***109***	***1753***



Supplemental Figure 1: Map of the states falling within each EPA region in the continental United States (adapted from U.S. Environmental Protection Agency, 2012).

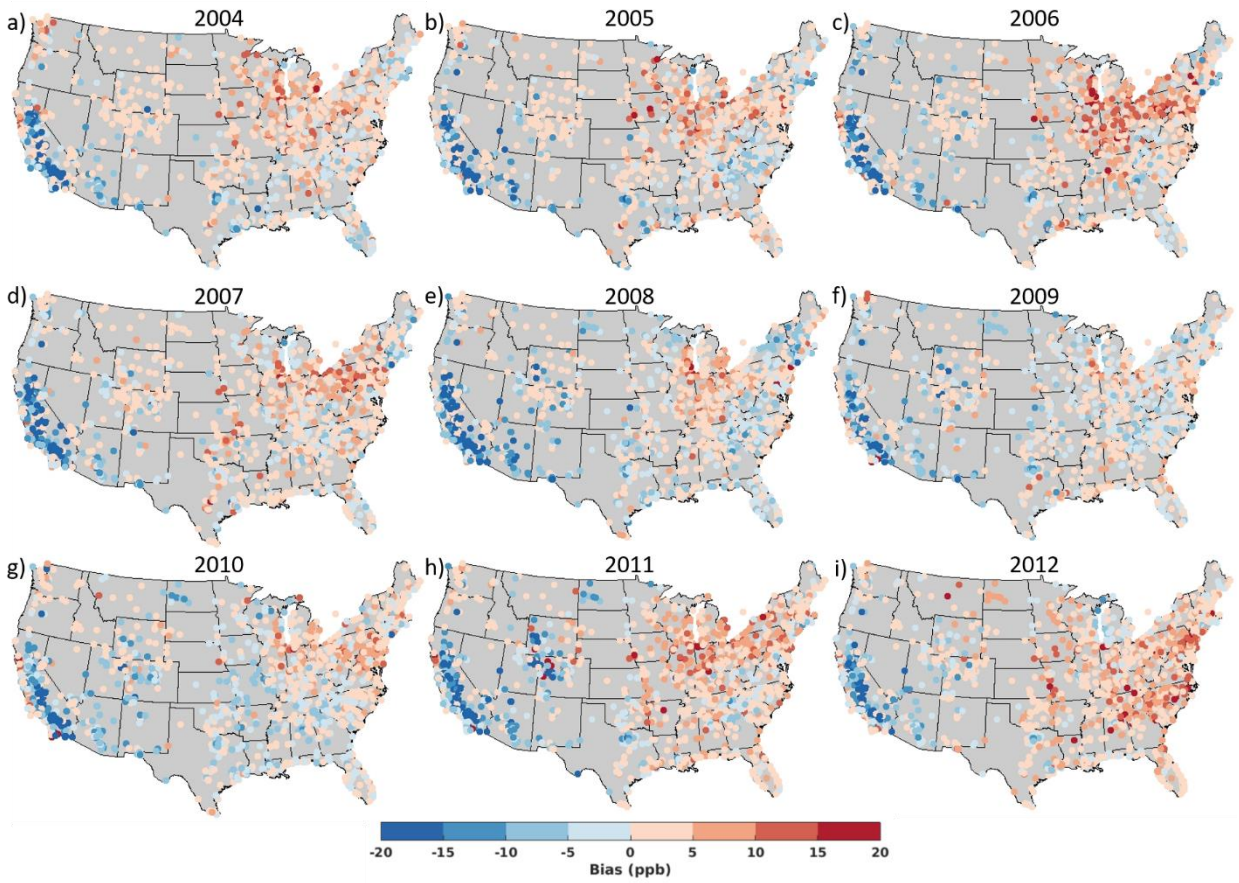
Supplemental Table 2: Number of EPA AQS sites collecting MDA8 O_3 data during each year from 2004-2012.

Number of EPA AQS Sites	
2004	1219
2005	1207
2006	1211
2007	1237
2008	1241
2009	1251
2010	1280
2011	1333
2012	1315

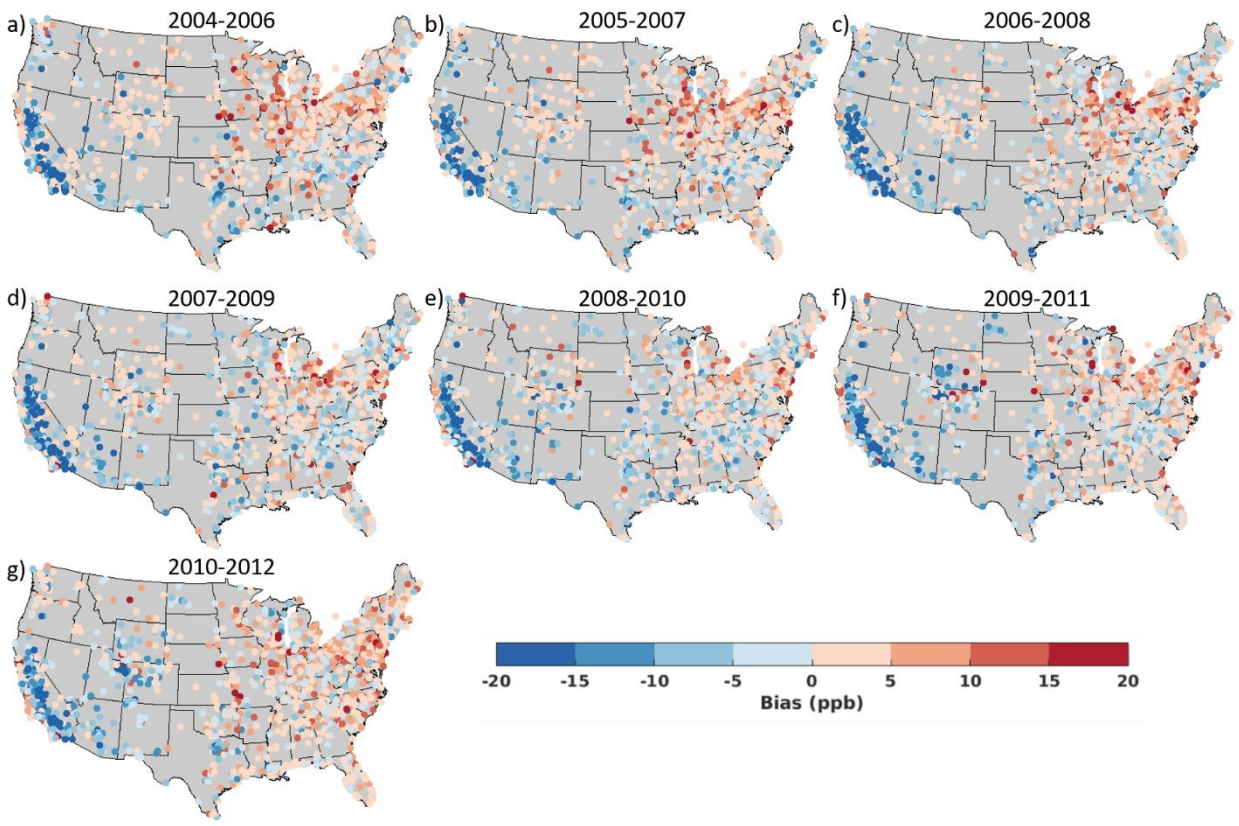
Supplemental Table 3: Global and US emissions totals for 2004-2012.

	Emissions	2004	2005	2006	2007	2008	2009	2010	2011	2012
Global	*Anthropogenic NO with biofuels (Tg N)*	30.3	30.2	30.1	29.9	29.5	29.0	28.8	28.8	28.8
	Biomass burning (Tg N)	4.5	4.7	4.6	4.6	3.9	3.5	5.0	3.7	3.7
	Soil (Tg N)	9.0	9.1	8.8	8.5	8.4	8.6	8.4	8.6	9.2
	Lightning (Tg N)	5.5	6.1	6.2	6.4	6.9	7.3	7.2	7.1	7.2
	Isoprene (Tg C)	493.0	499.3	471.5	453.6	435.3	455.4	466.0	453.3	467.3
US	*Anthropogenic NO with biofuels (Tg N)*	6.32	6.04	5.75	5.44	5.13	4.63	4.36	4.36	4.36
	Biomass burning (Tg N)	0.02	0.06	0.06	0.07	0.04	0.04	0.05	0.12	0.12
	Soil (Tg N)	0.78	0.86	1.02	0.92	0.82	0.79	0.77	0.95	1.10
	Lightning (Tg N)	0.86	0.86	0.77	0.75	1.10	1.13	1.13	1.28	1.31
	Isoprene (Tg C)	18.1	21.5	21.9	22.0	19.3	18.3	20.2	22.0	22.4

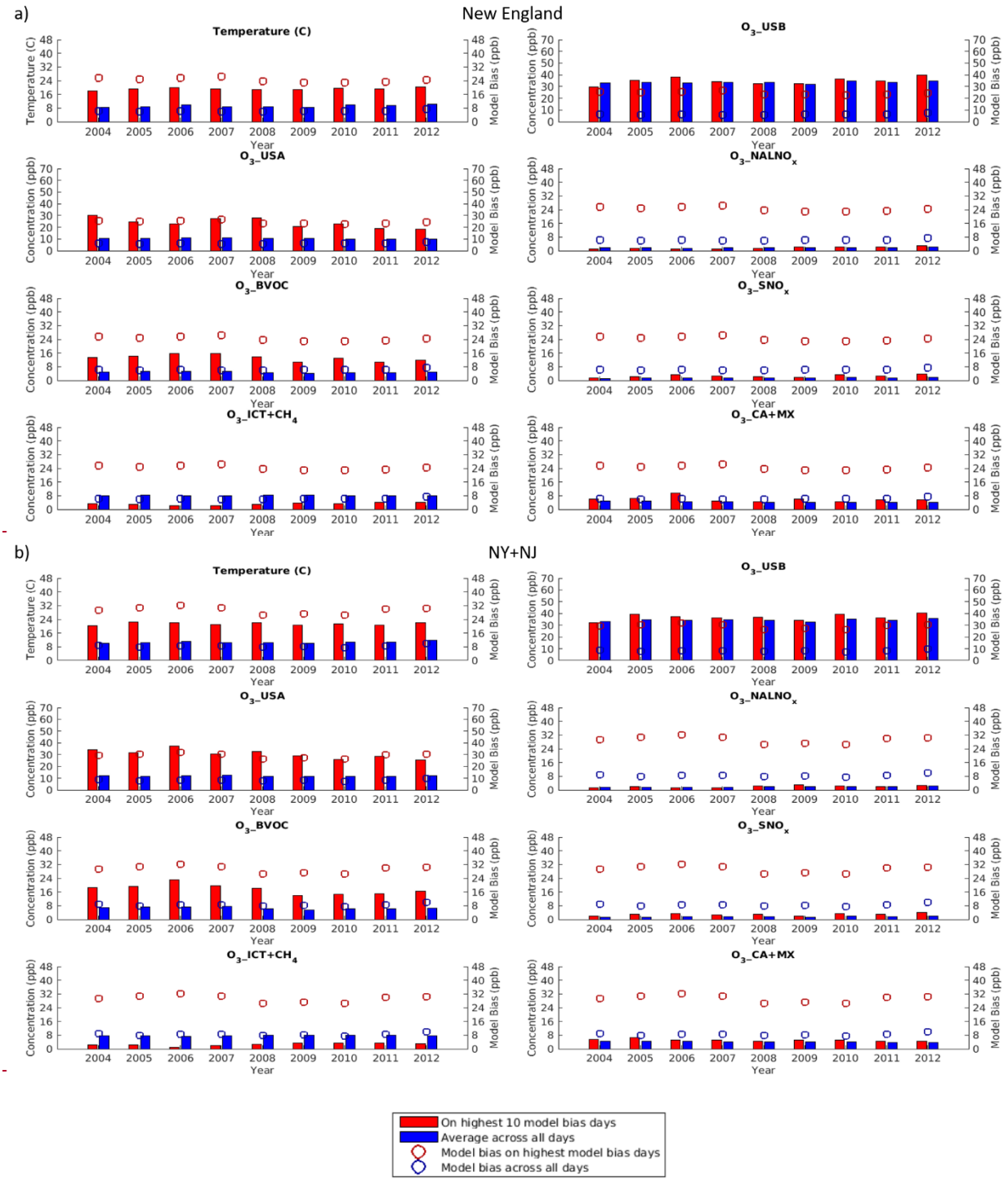
935



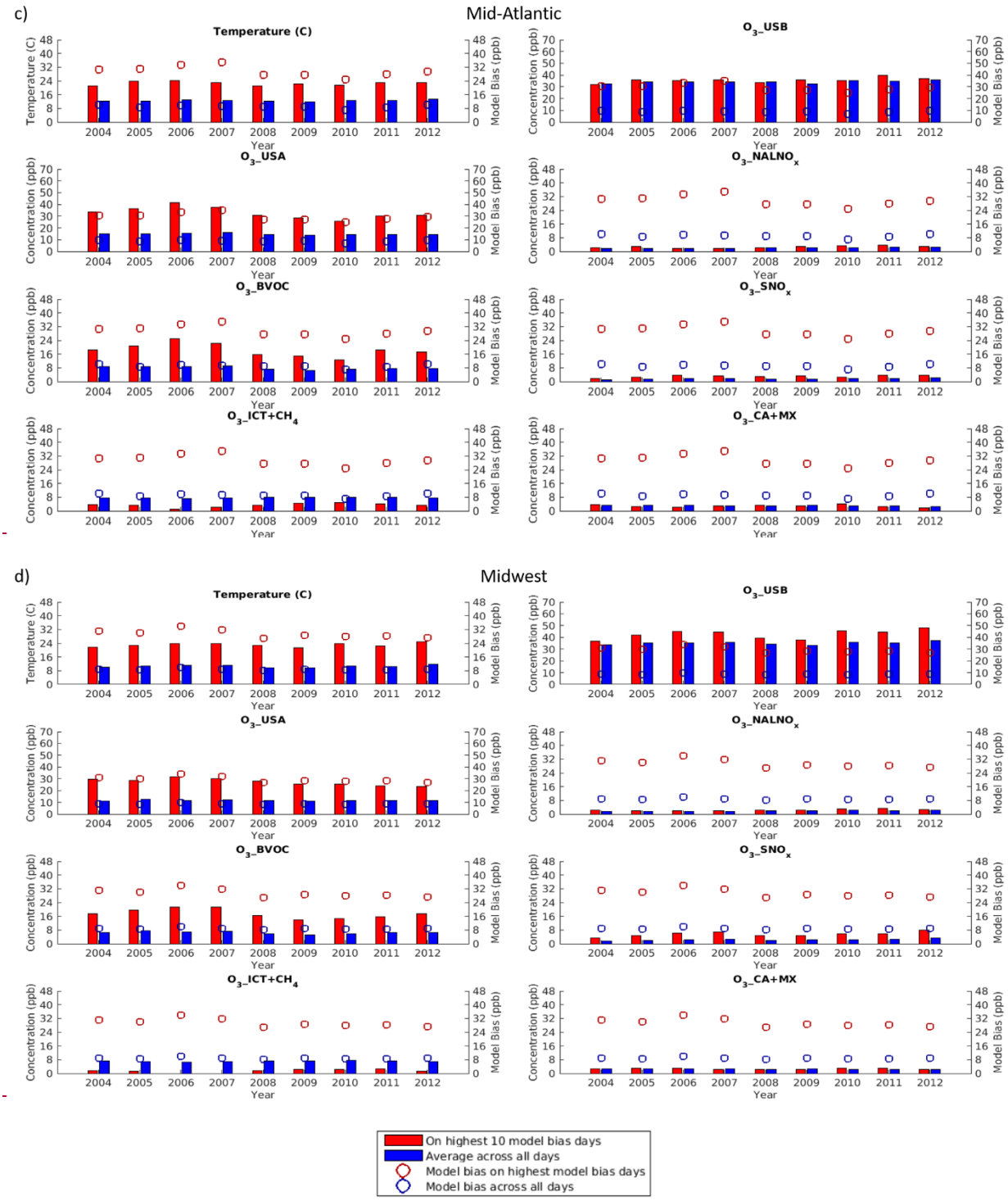
940 **Supplemental Figure 2: Average model bias (model – observed) on the O_3_top10obs days during (a) 2004, (b) 2005, (c) 2006, (d) 2007, (e) 2008, (f) 2009, (g) 2010, (h) 2011, and (i) 2012.**



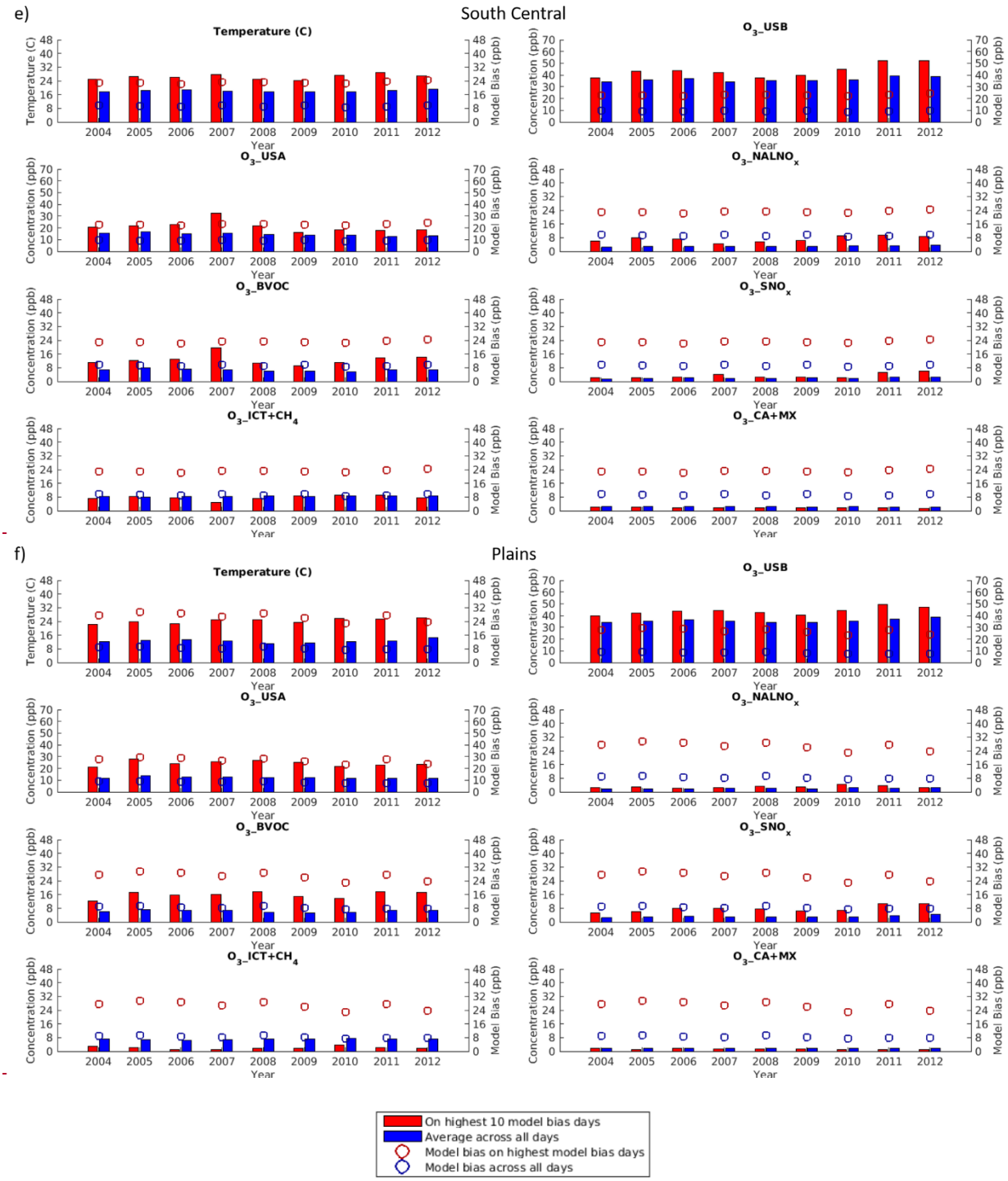
Supplemental Figure 3: Model bias (model – observed) on the 4th highest MDA8 O_3 day at each observational site averaged for each three-year span. (a) 2004-2006, (b) 2005-2007, (c) 2006-2008, (d) 2007-2009, (e) 2008-2010, (f) 2009-2011, and (g) 2010-2012.

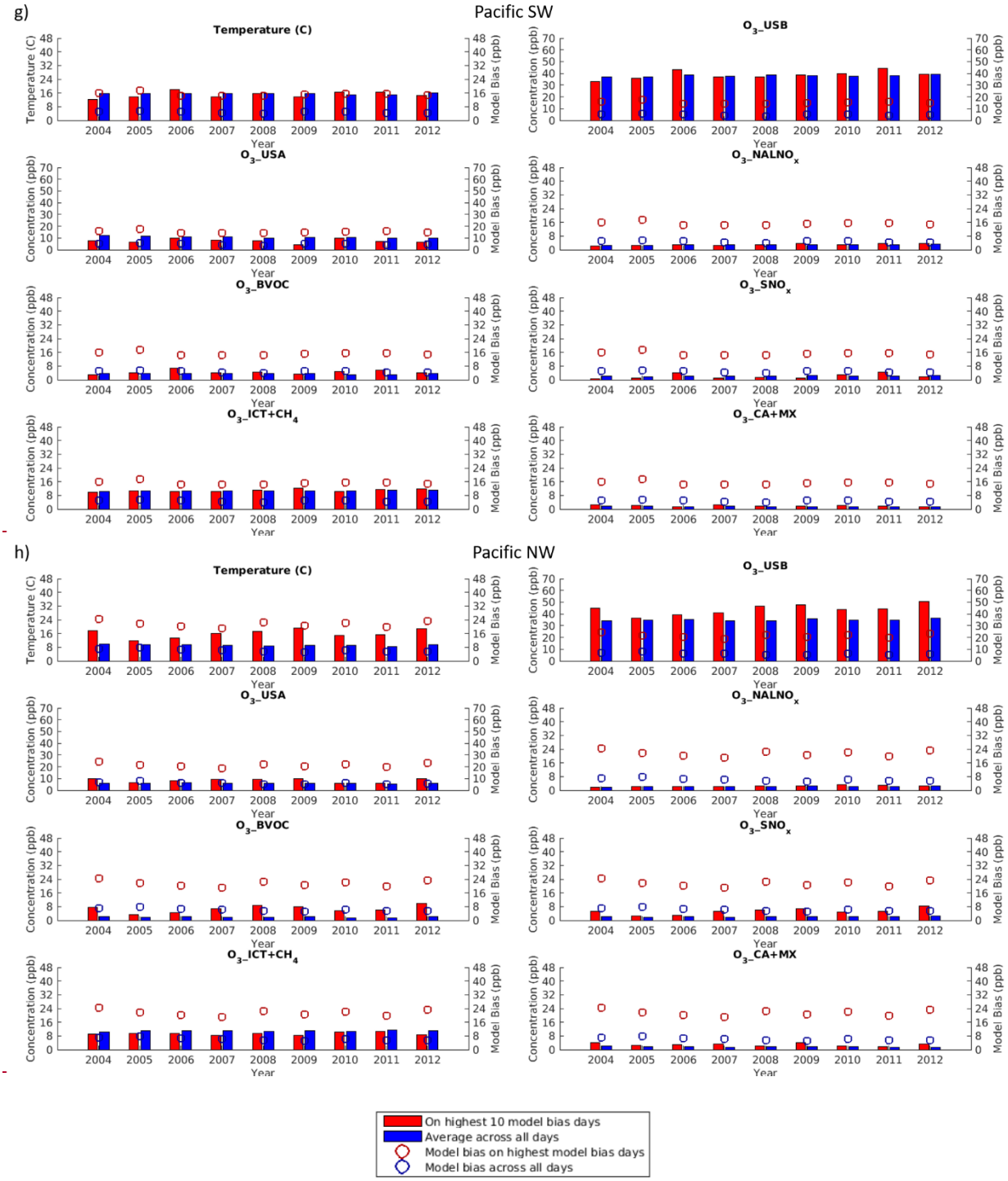

a)
New England
Temperature (C)
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
Temperature (C)
Concentration (ppb)
Model Bias (ppb)
Year
b)
NY+NJ
On highest 10 model bias days
Average across all days
Model bias on highest model bias days
Model bias across all days

945

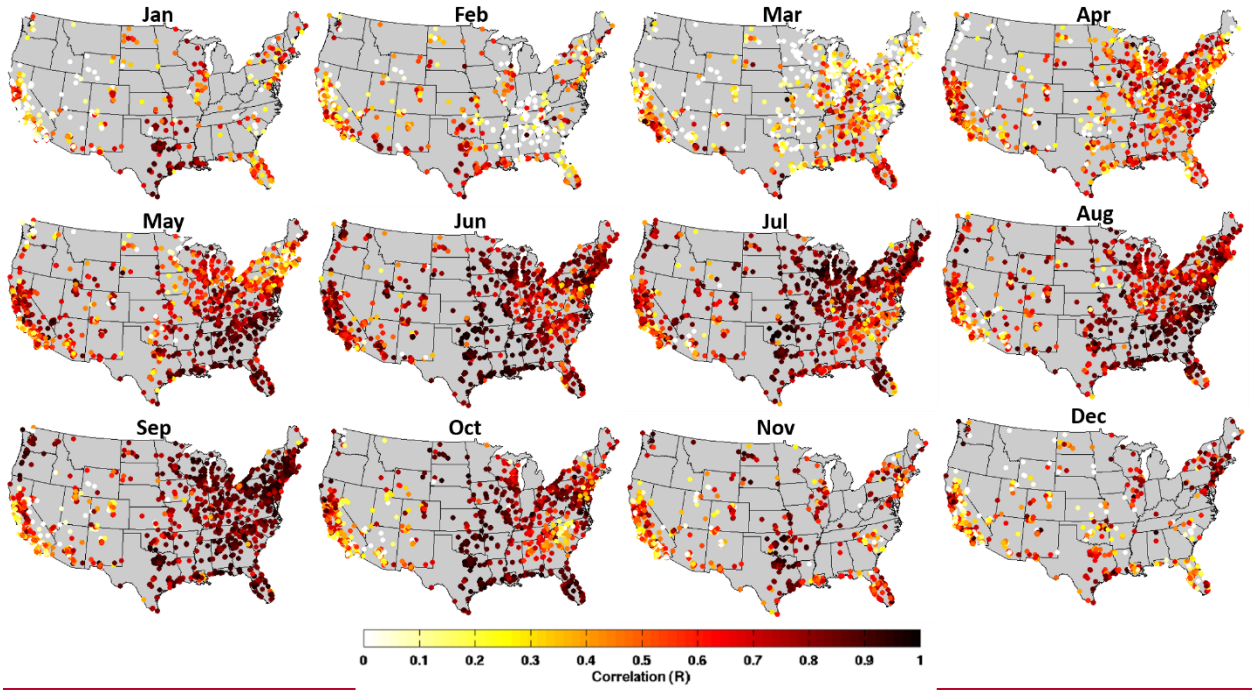

c)
Mid-Atlantic
Temperature (C)
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
d)
Midwest
Temperature (C)
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
Year
Model Bias (ppb)
Concentration (ppb)
On highest 10 model bias days
Average across all days
Model bias on highest model bias days
Model bias across all days

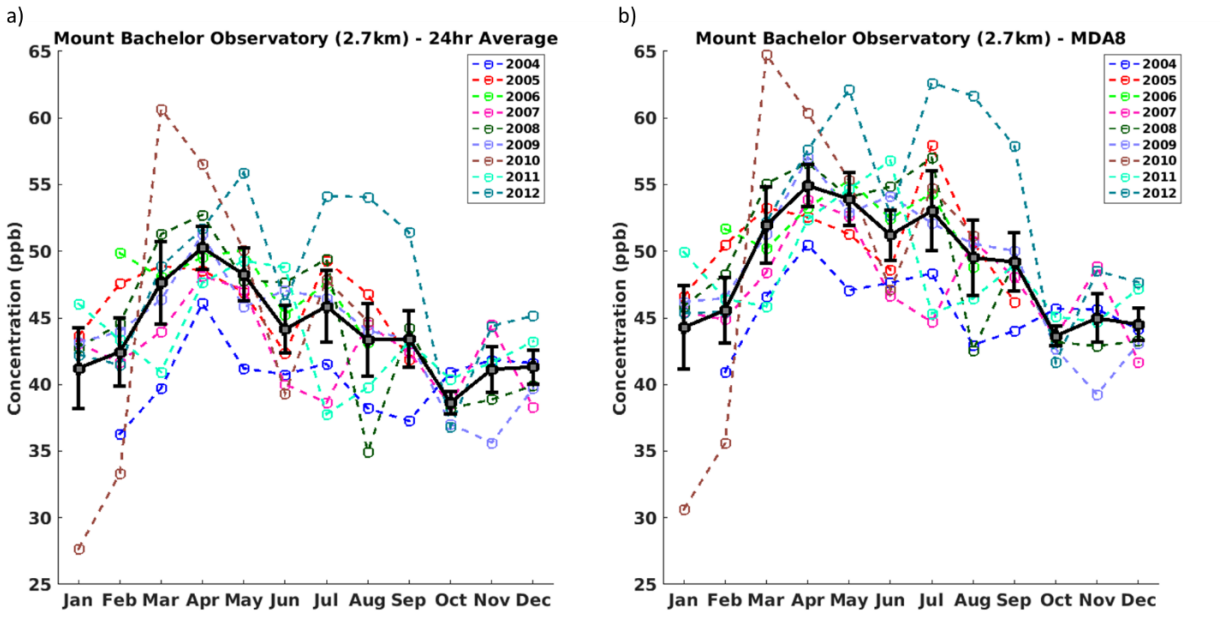
950


e)
South Central
Temperature (C)
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
f)
Plains
Temperature (C)
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
Temperature (C)
Concentration (ppb)
Model Bias (ppb)
Year
On highest 10 model bias days
Average across all days
Model bias on highest model bias days
Model bias across all days

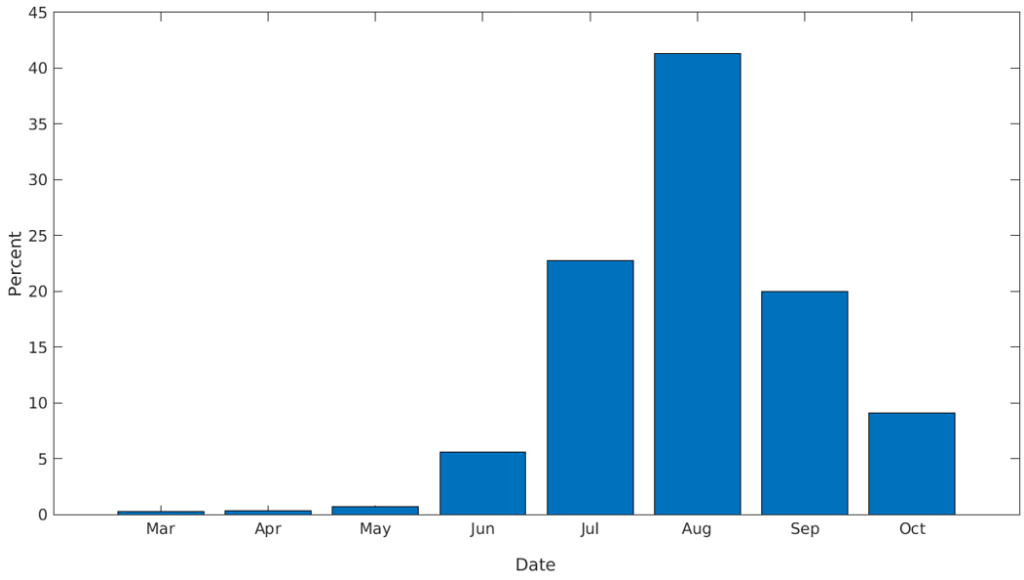

g)
Pacific SW
Temperature (C)
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
h)
Pacific NW
Temperature (C)
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
Year
Model Bias (ppb)
Concentration (ppb)
On highest 10 model bias days
Average across all days
Model bias on highest model bias days
Model bias across all days

955





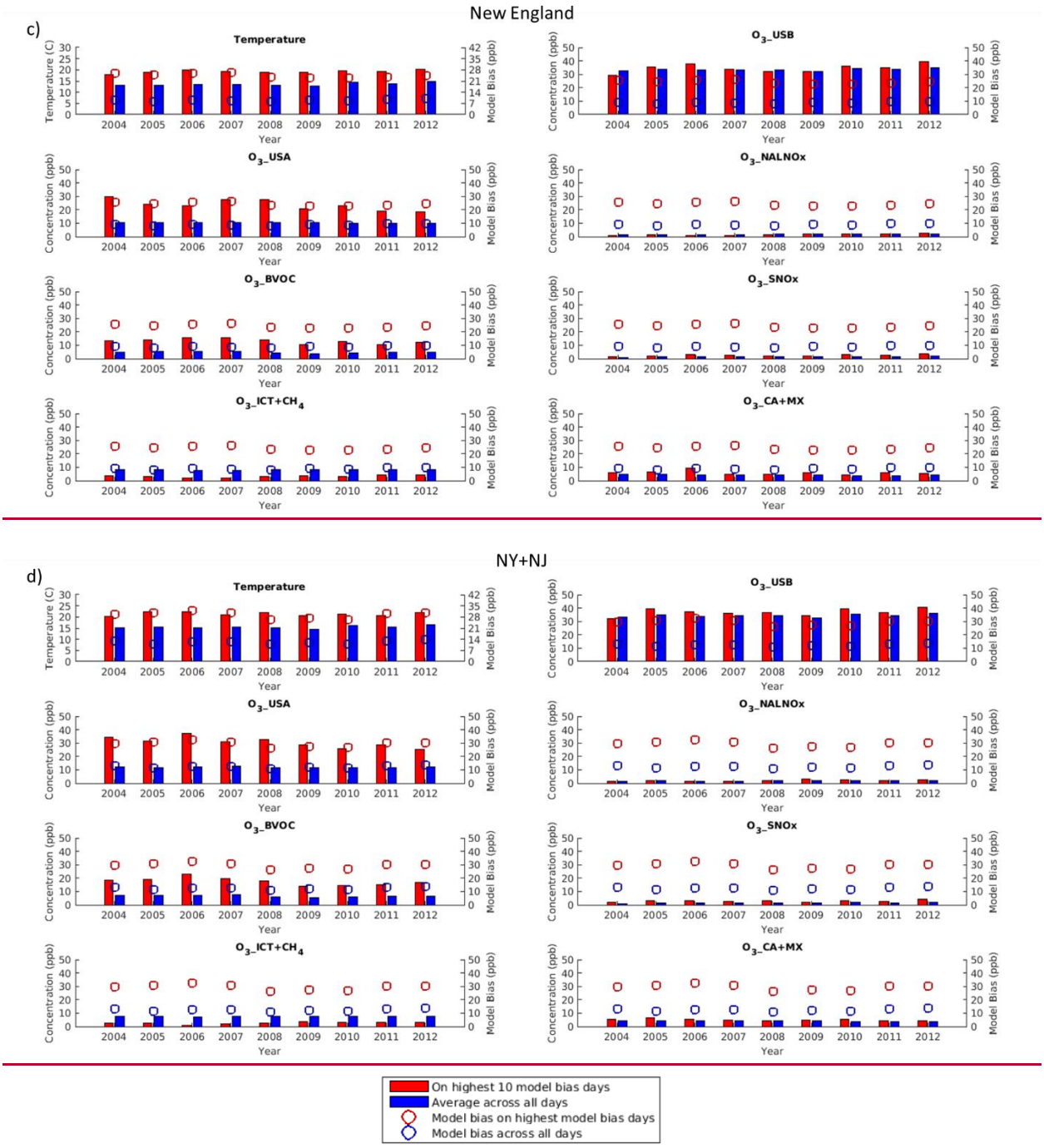
Supplemental Figure 4: Monthly average of observed (a) daily 24-hour and (b) MDA8 O_3 concentrations averaged across 2004-2012 at Mount Bachelor Observatory. Black line shows the average of each month from 2004-2012. Error bars show the standard deviation in the interannual variability in each month. Dashed lines show the concentrations for each individual year. ~~: Average influence of each sensitivity simulation on MDA8 O3 in each region on the 10 most biased days from Jan-Dec (red) versus averaged across all days (blue). Red circles show the average model bias (O3_Base – observations) on the top 10 model bias days. Blue circles show the model bias averaged across all days. The circles do not vary between subplots. Note that O3_USB and O3_USA are on a different scale than the other plots.~~

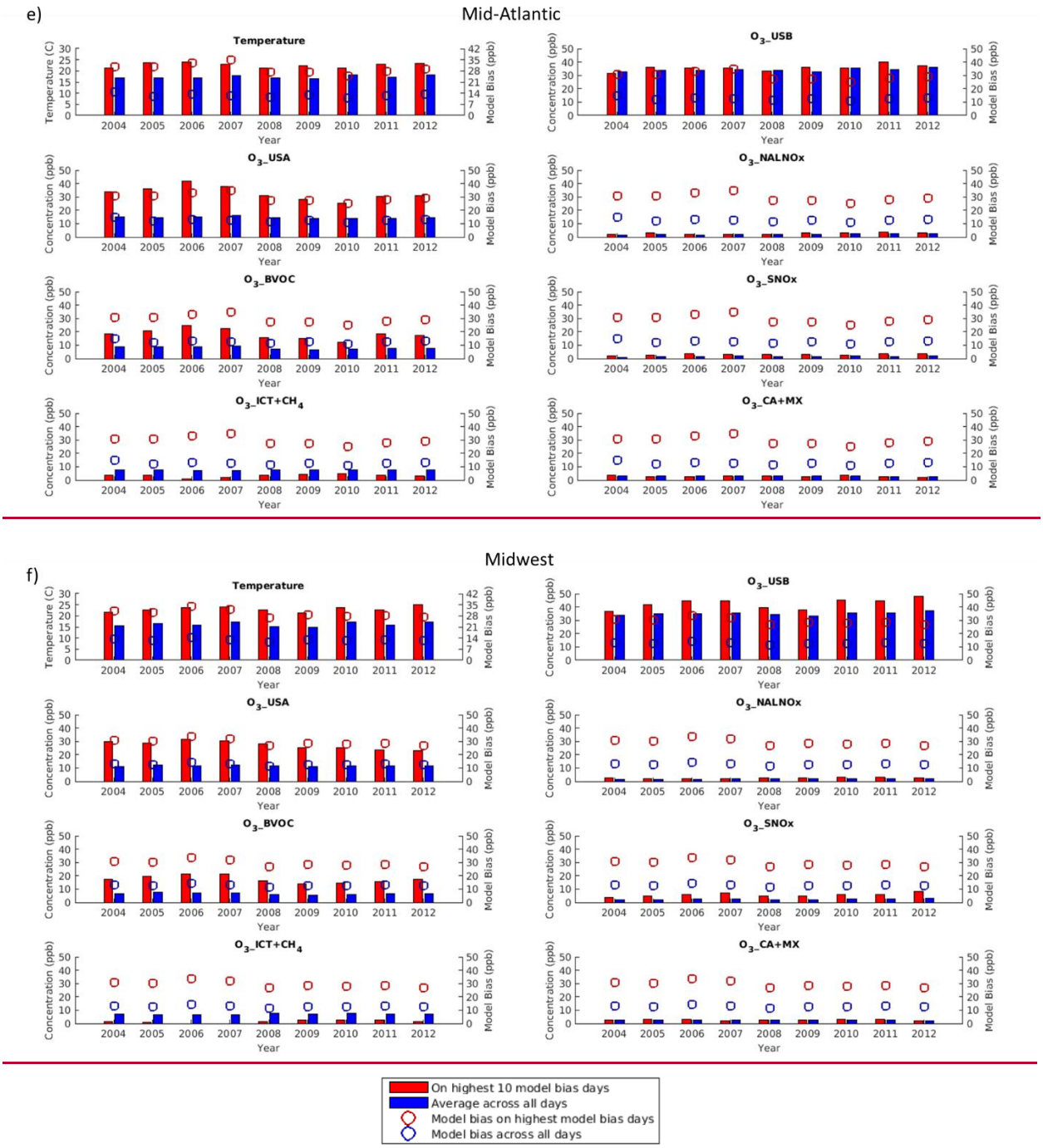


965

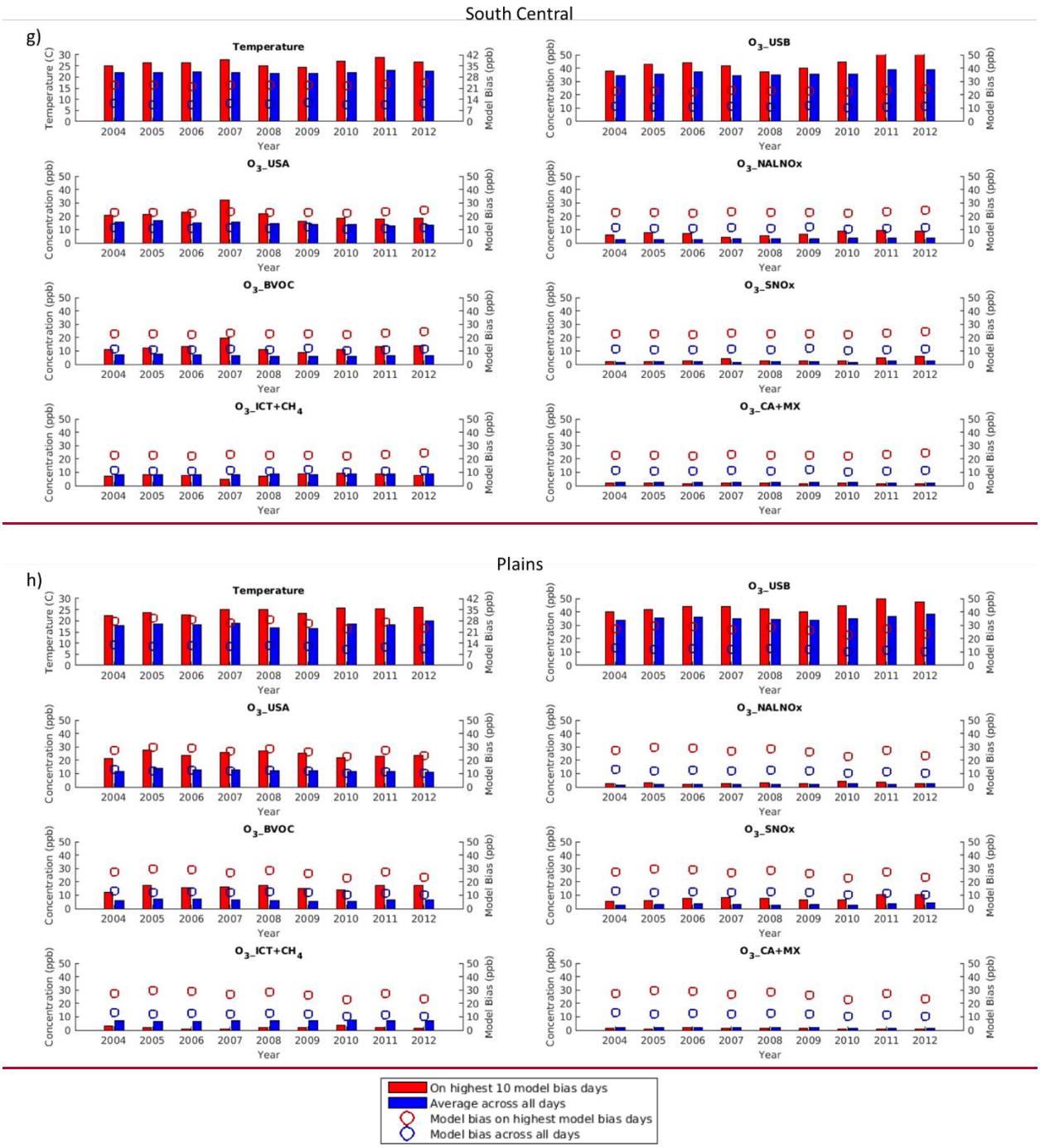
Supplemental Figure 5: Percent of total top 10 most biased days from Jan-Dec (9 years x 10 days x 10 regions) that fell within each month in the United States. All the most biased days fell between Mar-Oct.

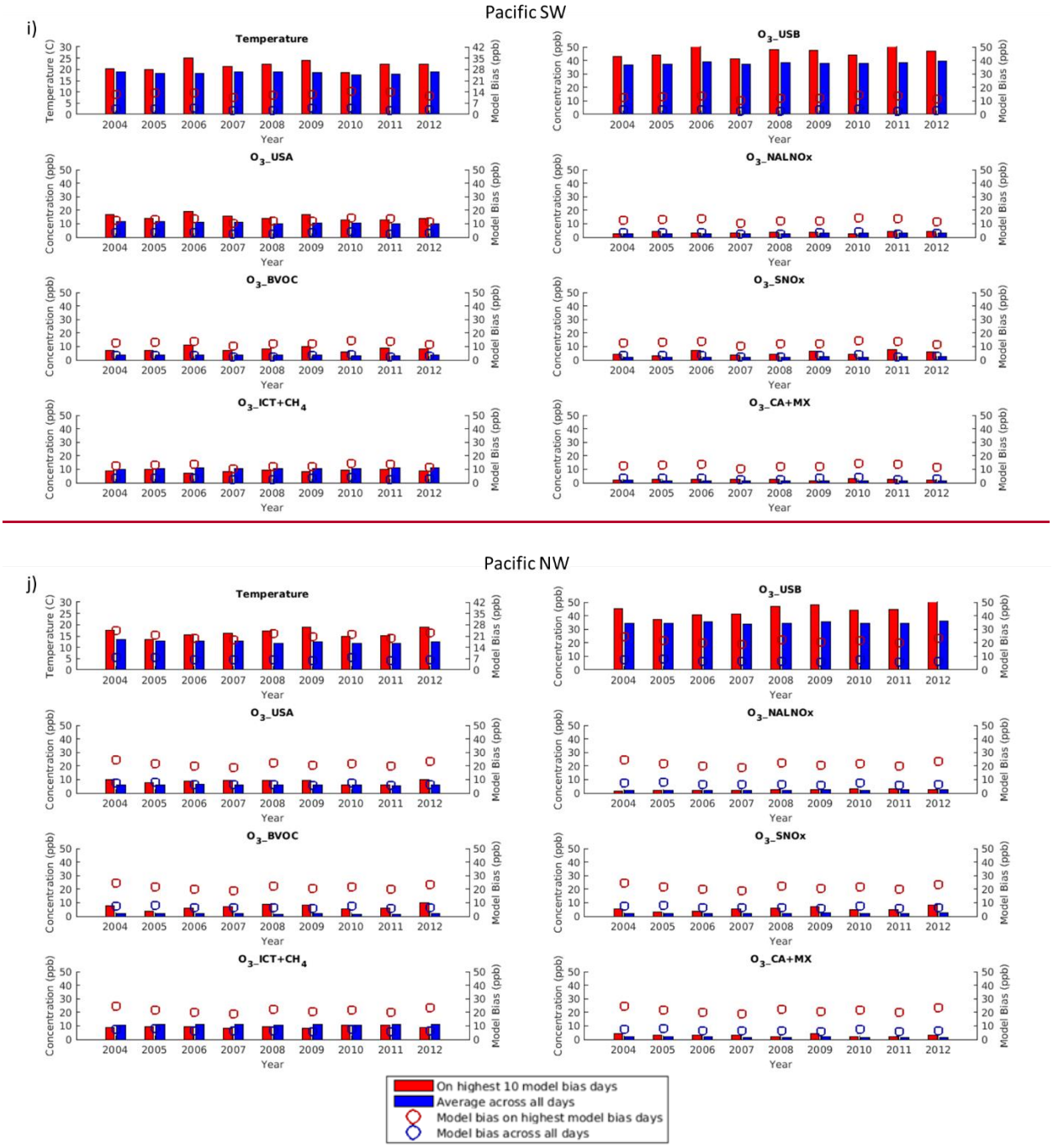

a)
Southeast
Temperature
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
Temperature (C)
Concentration (ppb)
Model Bias (ppb)
Year
b)
Mountain and Plains
On highest 10 model bias days
Average across all days
Model bias on highest model bias days
Model bias across all days


New England
c)
Temperature
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
Temperature (C)
Concentration (ppb)
Model Bias (ppb)
Year
NY+NJ
d)
On highest 10 model bias days
Average across all days
Model bias on highest model bias days
Model bias across all days

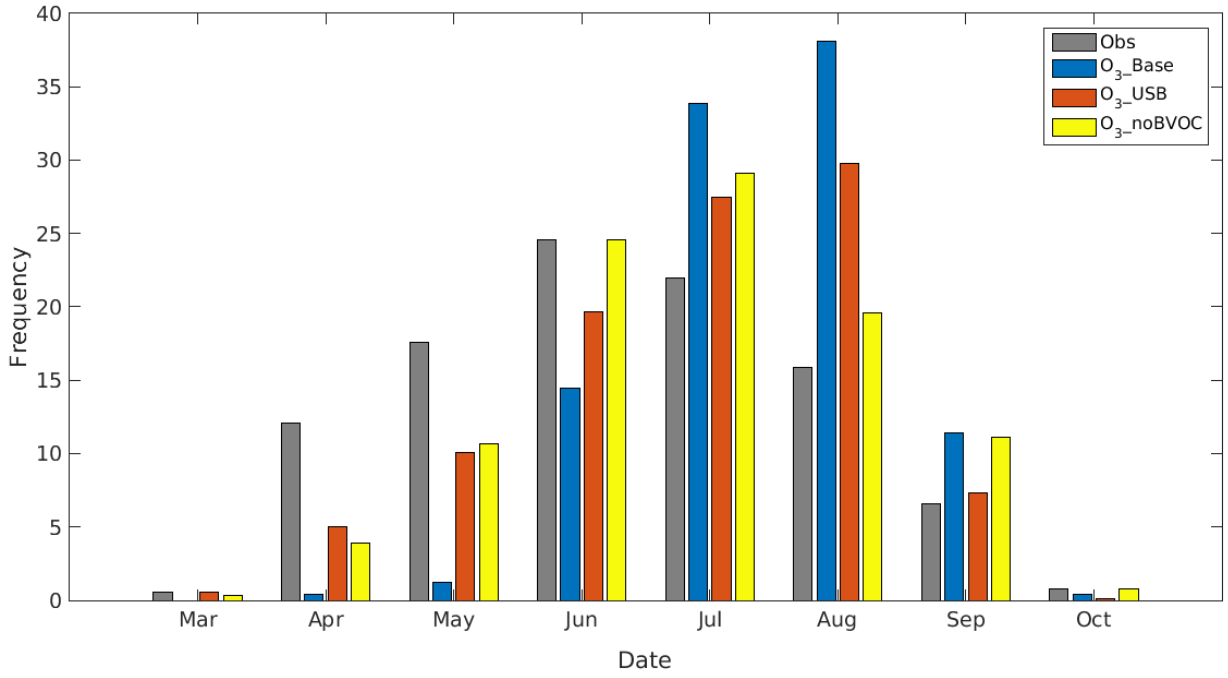

e)
Mid-Atlantic
Temperature
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
Temperature (C)
Concentration (ppb)
Model Bias (ppb)
Year
f)
Midwest
On highest 10 model bias days
Average across all days
Model bias on highest model bias days
Model bias across all days

980

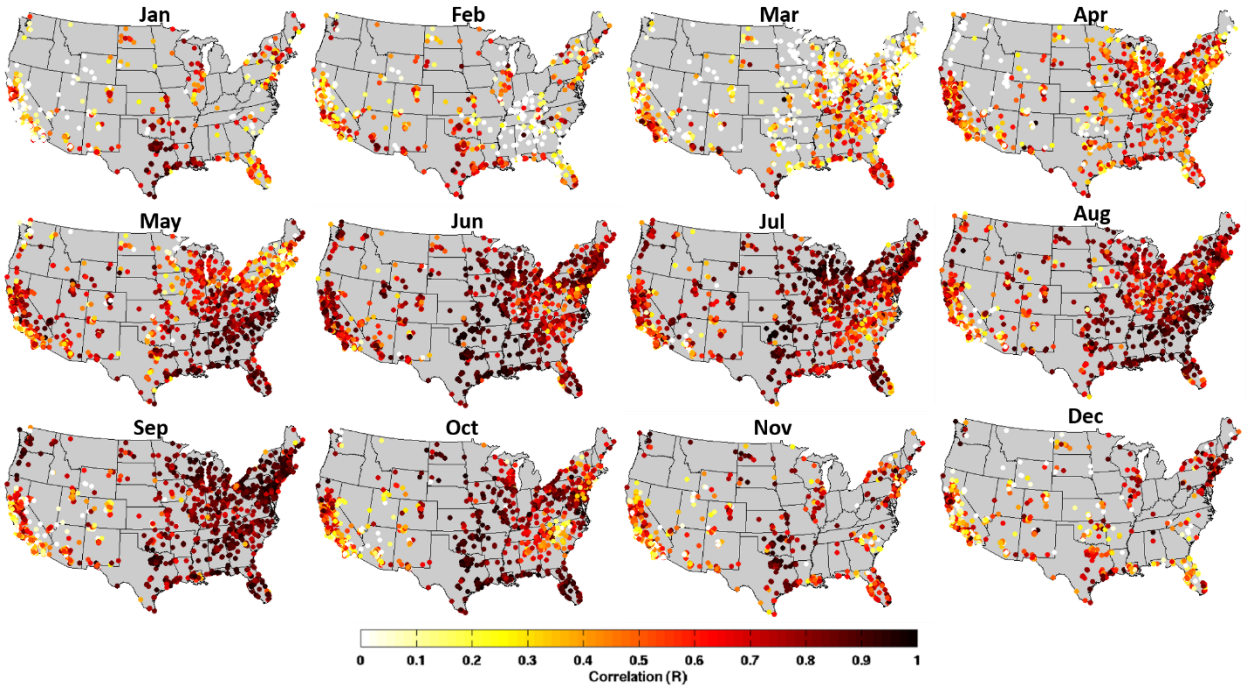

South Central
g)
Temperature
O3_USB
O3_USA
O3_NALNOx
O3_BVOC
O3_SNOx
O3_ICT+CH4
O3_CA+MX
Temperature (C)
Concentration (ppb)
Model Bias (ppb)
Year
Plains
h)
On highest 10 model bias days
Average across all days
Model bias on highest model bias days
Model bias across all days



Supplemental Figure 6: Average influence of each sensitivity simulation on MDA8 O_3 in each region on the 10 most biased days from Mar-Oct (red) versus averaged across all days (blue). Red circles show the average model bias (O_3_Base – observations) on the top 10 model bias days. Blue circles show the model bias averaged across all days. The circles do not vary between subplots.

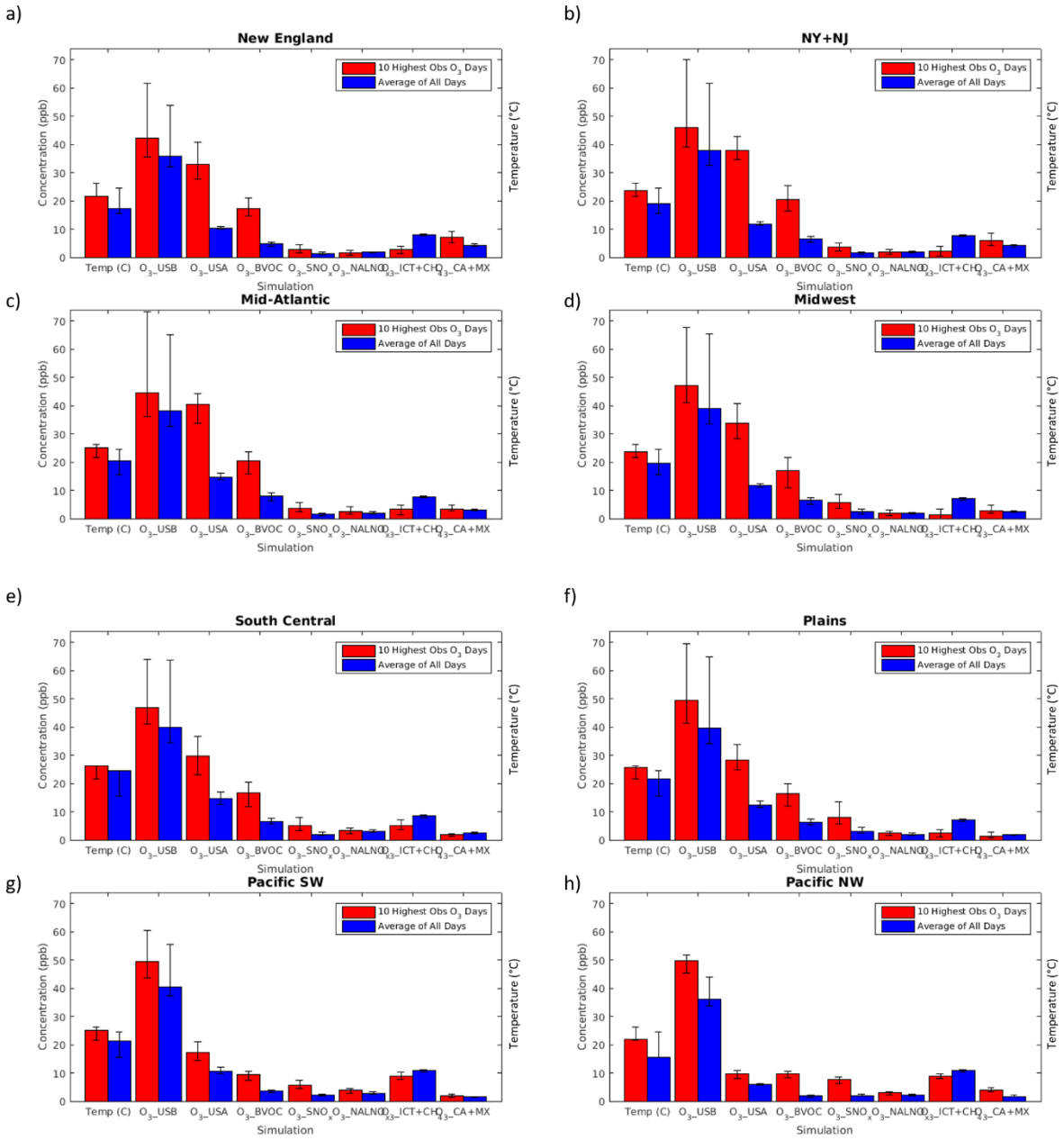


Supplemental Figure 7: Percent of total top ten days (9 years x 10 days x 10 regions) from Jan-Dec (365 or 366 days) in the observations, O_3_Base, O_3_USB, and O_3_noBVOC that fell within each month for all sites across the U.S.A. All the top ten days for each simulation fell between Mar-Oct.

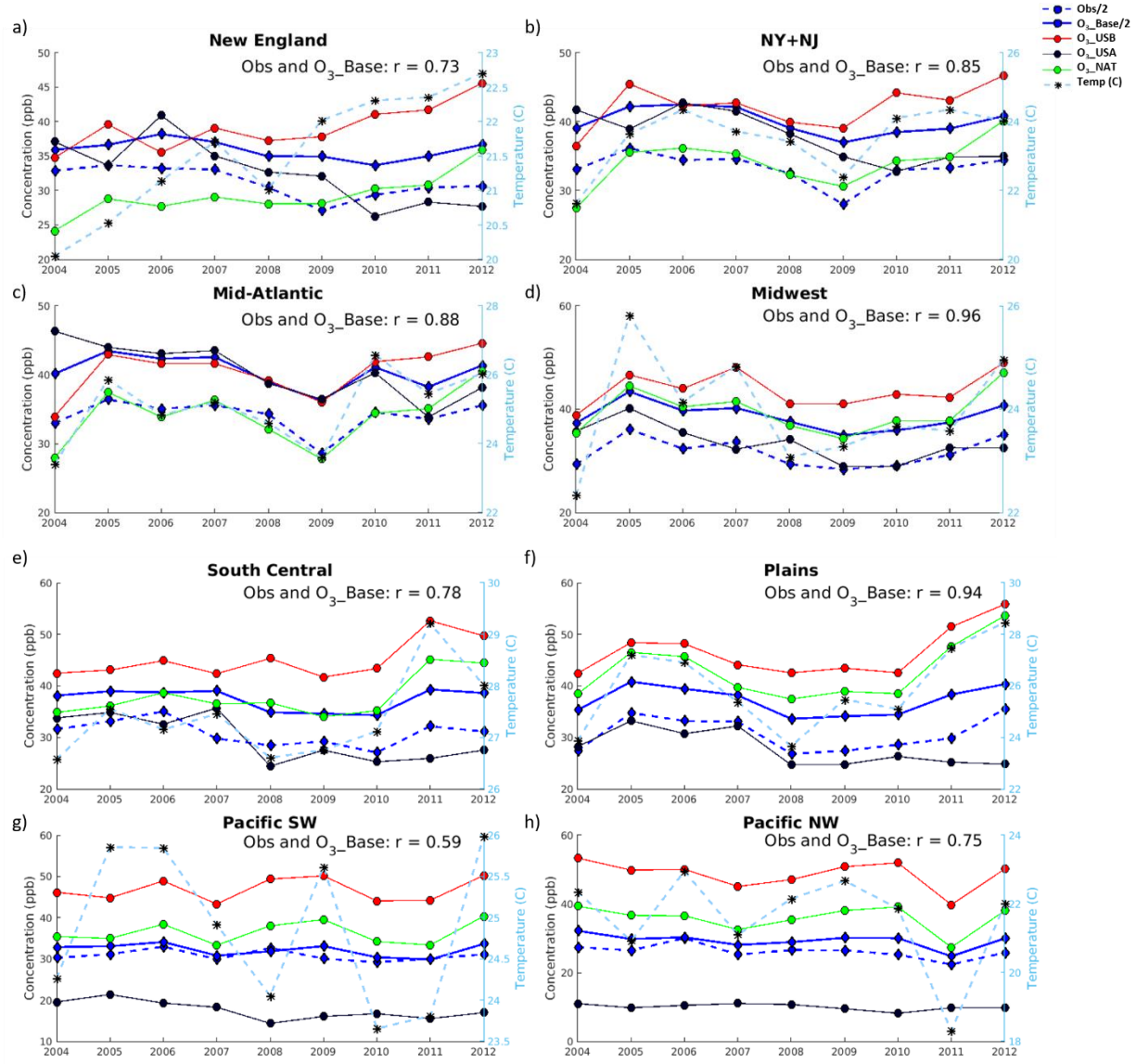


Supplemental Figure 8: Correlation between 2004-2012 year-to-year monthly averages for MDA8 O_3 in the observation and in the model (O_3_Base) for each individual month.

995



1000 **Supplemental Figure 9: Average 2004-2012 influence of each sensitivity simulation to O3_Base in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW on the MDA8 O3_top10obs_JJA days (red) versus averaged across all days (blue). Error bars show the average concentration on the lowest versus highest year for each sensitivity simulation in each region.**



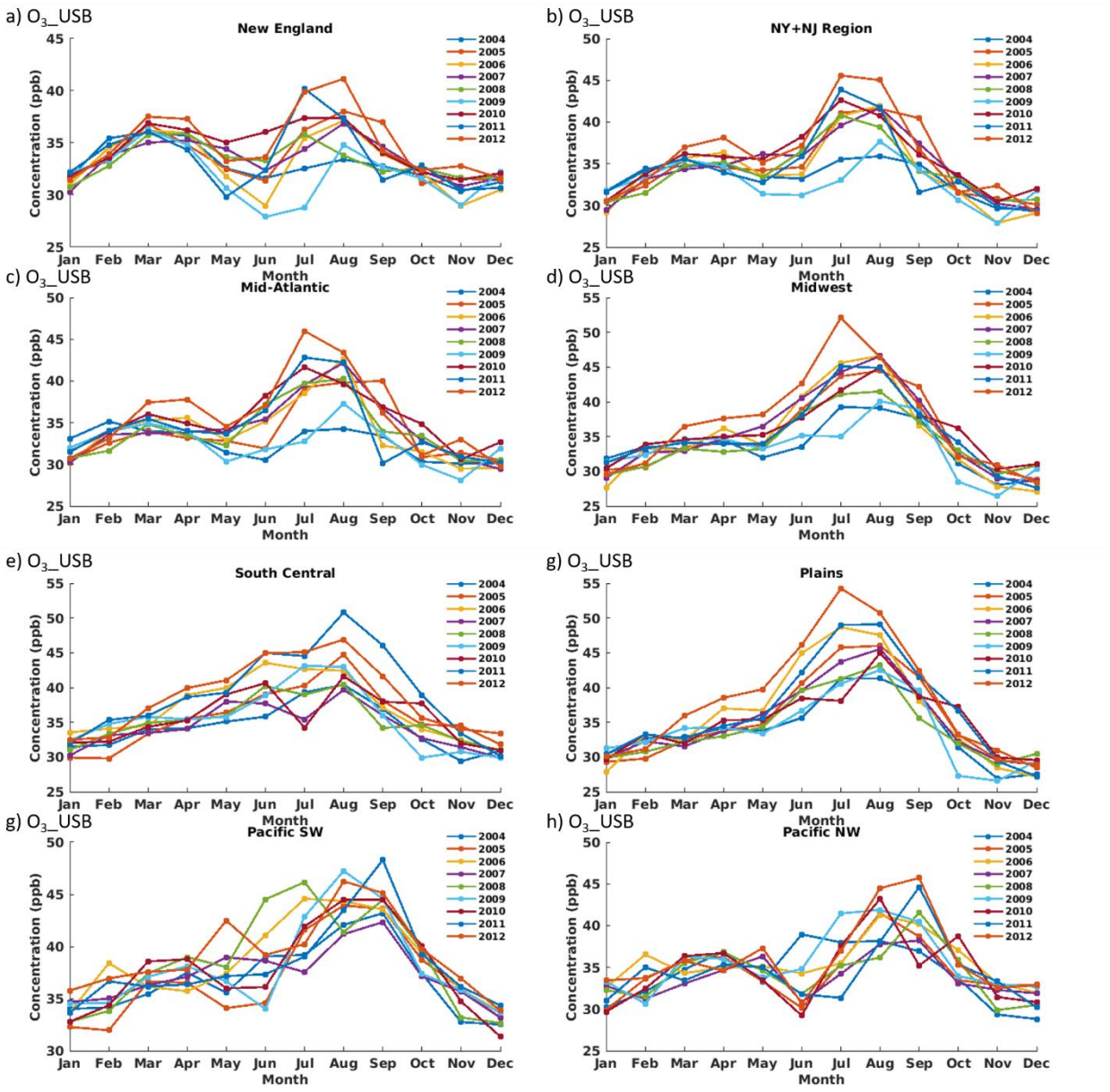
Supplemental Figure 10: Average yearly MDA8 O_3_top10obs_JJA concentrations for observations (divided by 2 to fit on the same axes; blue dashed line), O_3_Base (divided by 2; blue solid line), O_3_USB (red), O_3_USA (black), O_3_NAT (green) MDA8, and temperature (in degrees C; light blue) sampled on the O_3_top10obs days in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW.

Supplemental Table 4: Monthly average temperature across all days in each season (average of 2004-2012) in (1) GEOS-Chem, in (2) the Global Historical Climatology Network (GHCN) and the Climate Anomaly Monitoring System (CAMS) (in degrees C), and (3) the difference between these values.

	Model Temperature (C)			GHCN+CAMS Temperature (C)			Model Temp. Bias		
Region	*MAM*	*JJA*	*SON*	*MAM*	*JJA*	*SON*	*MAM*	*JJA*	*SON*
New England	7	19	11	8	20	11	0	-1	1
NY+NJ	9	21	12	9	21	12	0	0	1
Mid-Atlantic	12	23	14	12	23	14	0	0	0
Southeast	17	26	18	17	26	18	0	0	0
Midwest	10	22	12	10	22	12	0	0	0
South Central	18	27	19	19	28	20	-1	-1	-1
Plains	13	25	13	13	25	13	0	0	0
Mountains + Plains	7	20	9	7	19	8	0	1	1
Pacific SW	14	23	17	14	22	17	-1	0	0
Pacific NW	8	17	10	8	17	9	0	0	1

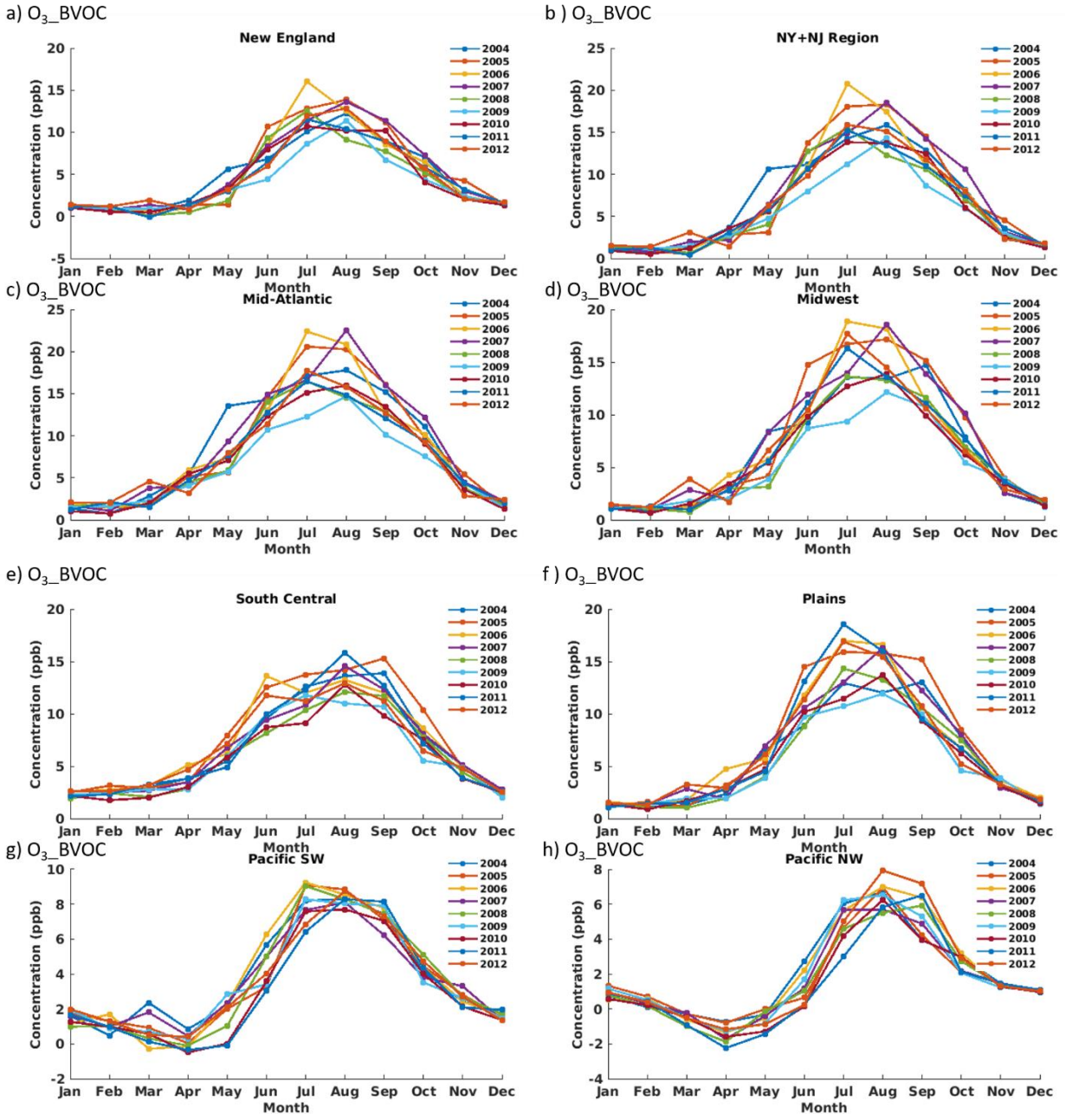
Supplemental Table 5: Correlation between (1) O_3_Base and O_3_USB and (2) O_3_Base and O_3_USA on the average of O_3_top10obs_JJA days from 2004-2012 in each region.

	Correlation	
Region	O_3_Base and O_3_USB	O_3_Base and O_3_USA
New England	0.28	0.64
NY+NJ	0.50	0.58
Mid-Atlantic	0.54	0.70
Southeast	0.66	0.59
Midwest	0.75	0.76
South Central	0.71	0.72
Plains	0.80	0.75
Mountains + Plains	0.95	0.64
Pacific SW	0.72	0.28
Pacific NW	0.98	0.05



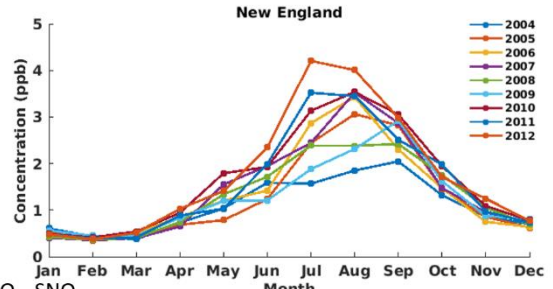
1020

Supplemental Figure 11: Monthly average MDA8 O_3_USB concentrations in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW. ~~**Supplemental Figure 8: Monthly average MDA8 O_3_BVOC**~~

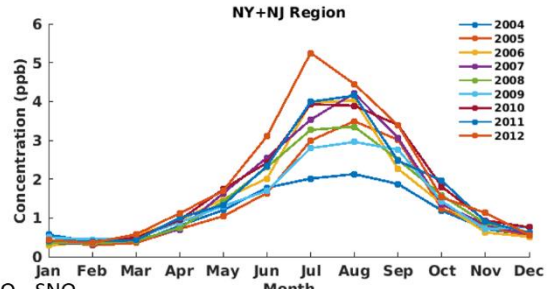

a) O3_BVOC
New England
b) O3_BVOC
NY+NJ Region
c) O3_BVOC
Mid-Atlantic
d) O3_BVOC
Midwest
e) O3_BVOC
South Central
f) O3_BVOC
Plains
g) O3_BVOC
Pacific SW
h) O3_BVOC
Pacific NW
Concentration (ppb)
Month
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2004 2005 2006 2007 2008 2009 2010 2011 2012

1025

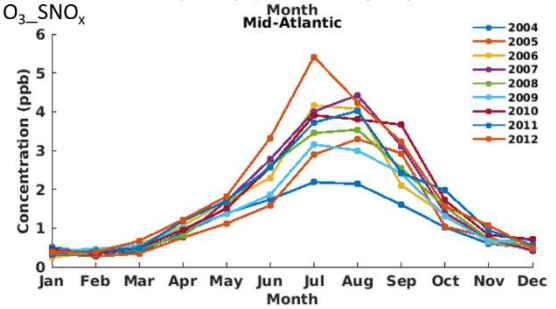
a) O_3_SNO_x



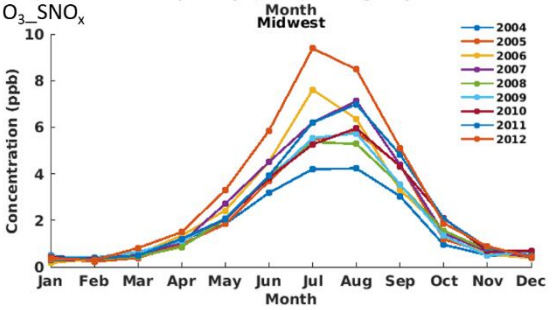
b) O_3_SNO_x



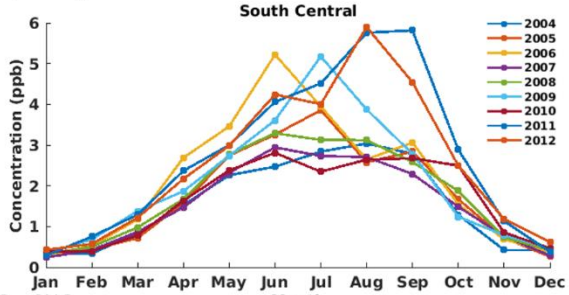
c) O_3_SNO_x



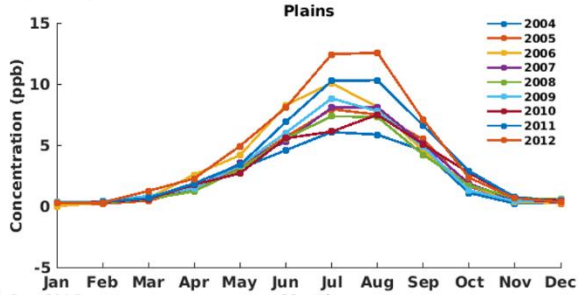
d) O_3_SNO_x



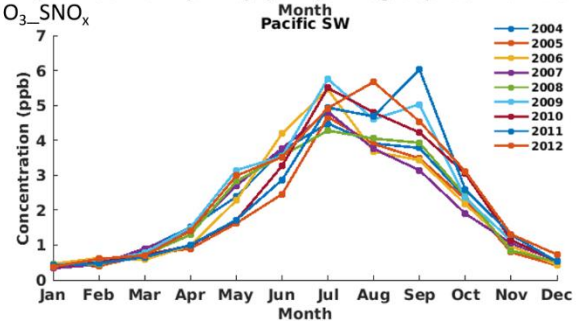
e) O_3_SNO_x



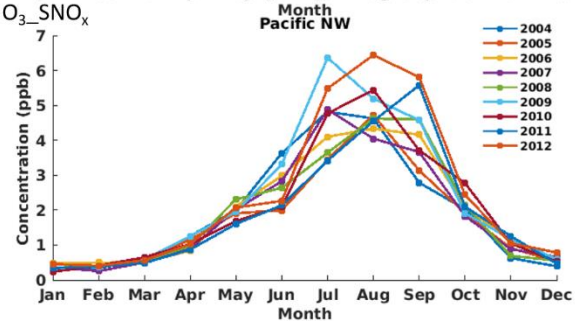
f) O_3_SNO_x

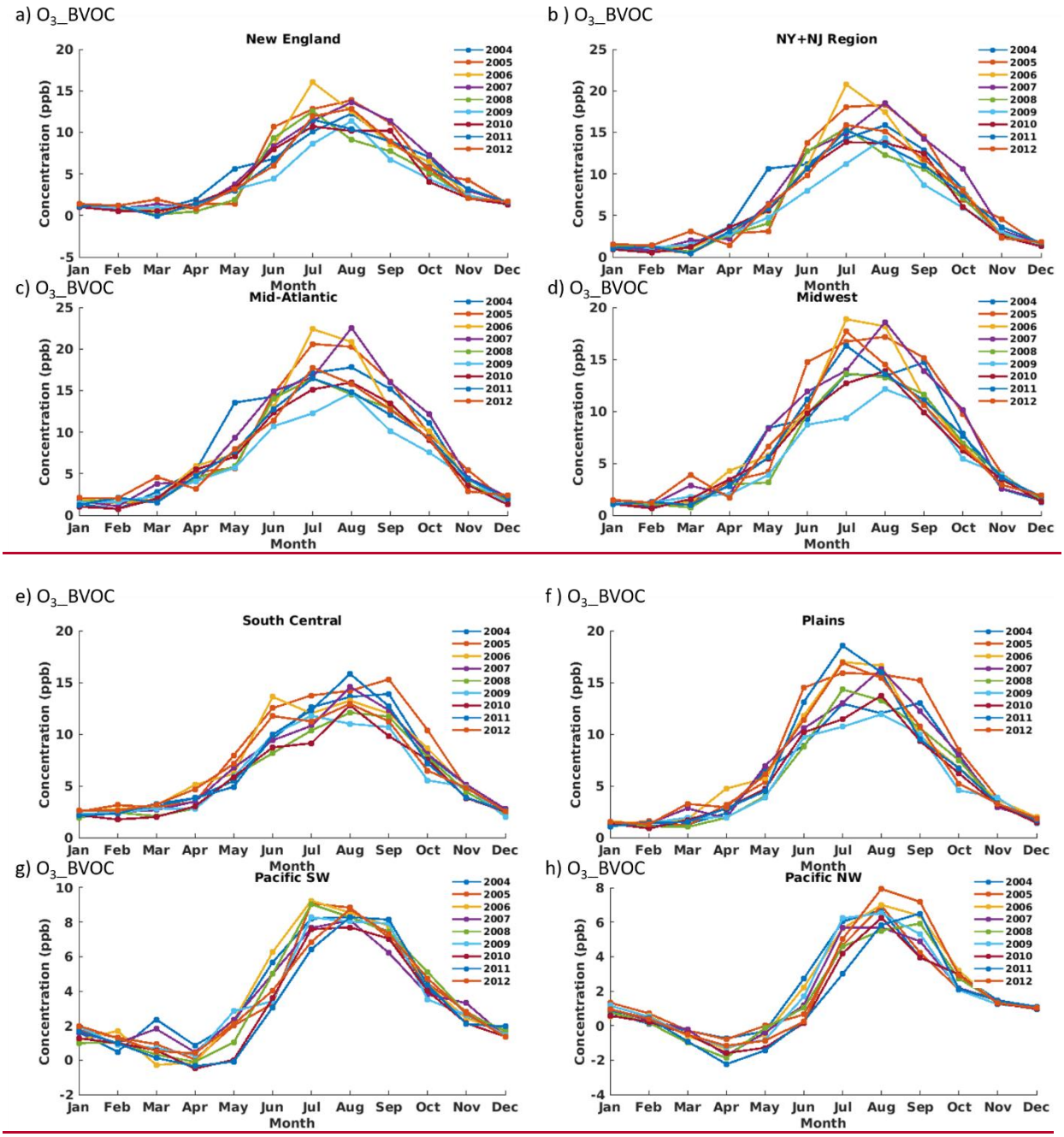


g) O_3_SNO_x

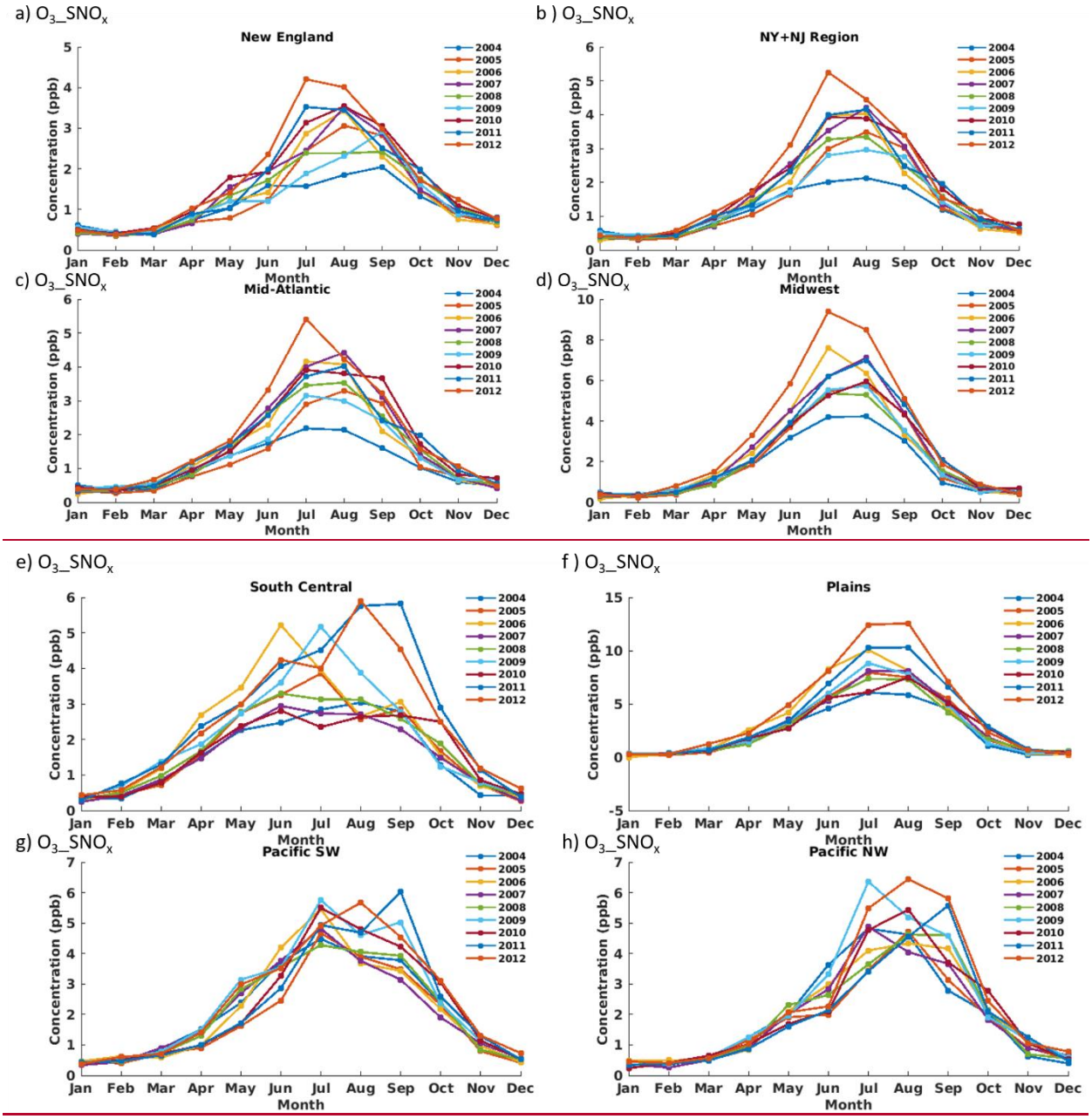


h) O_3_SNO_x

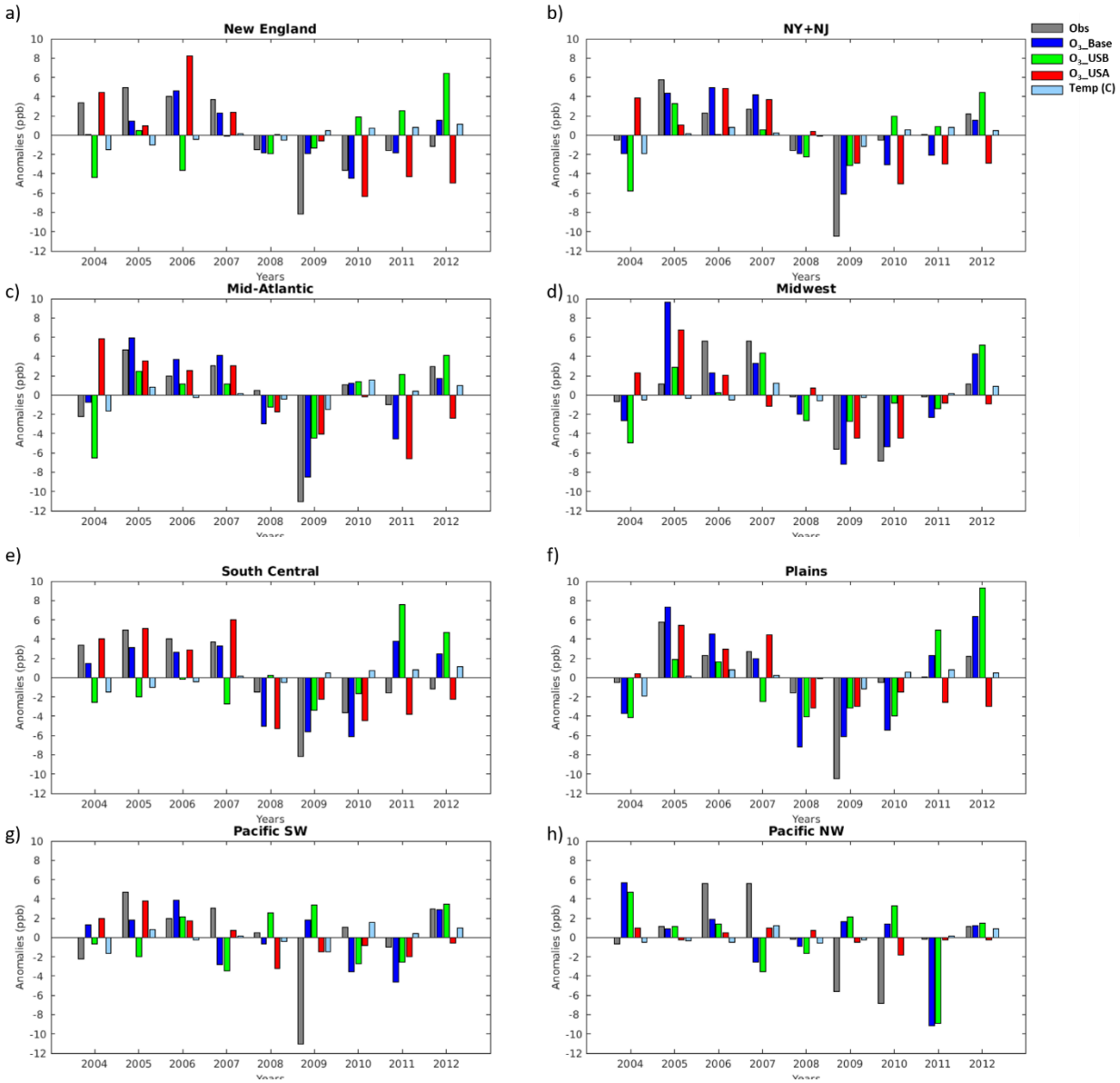




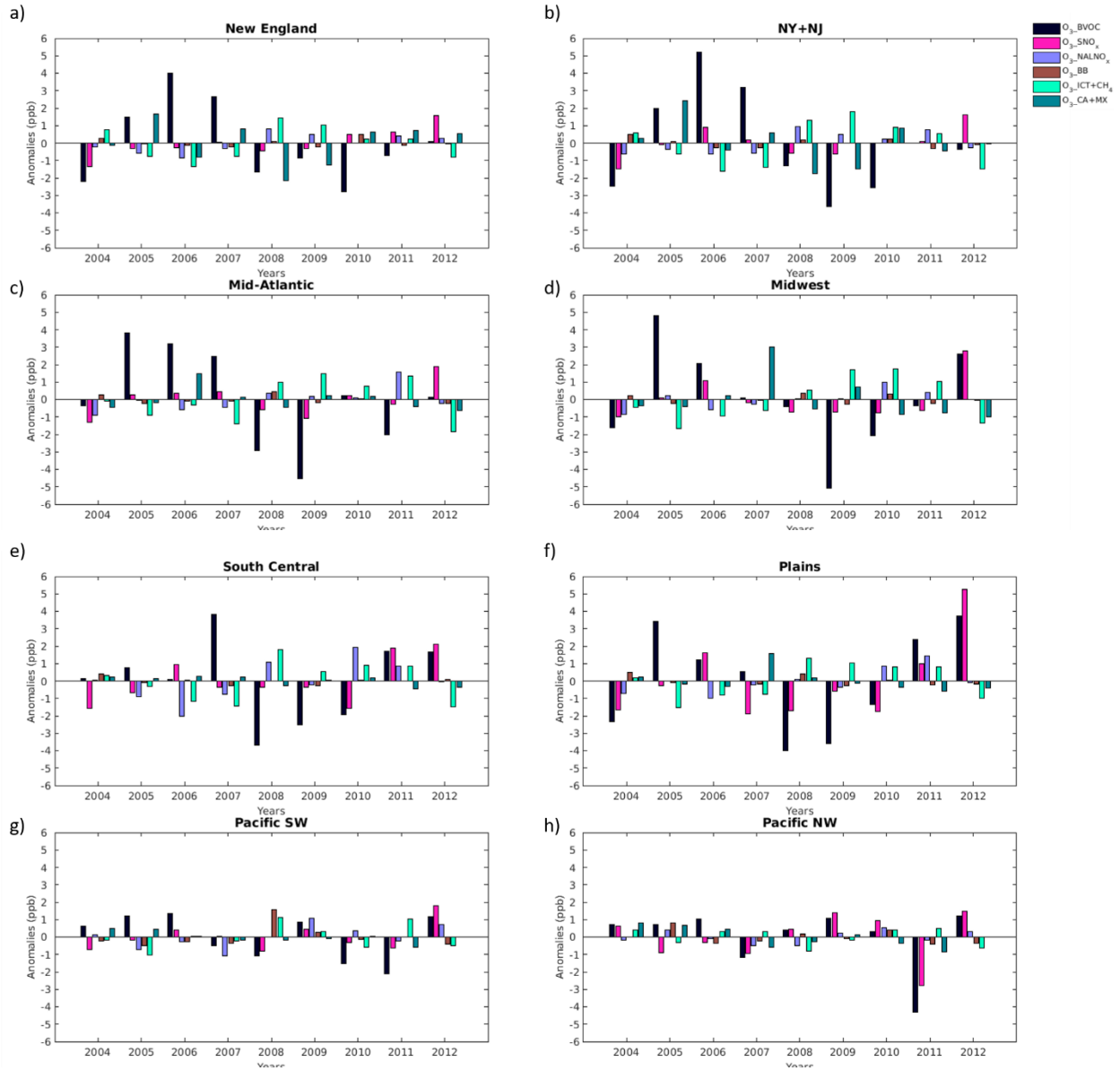
Supplemental Figure 12: Monthly average MDA8 O_3 BVOC concentrations in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW.



Supplemental Figure 13: Monthly average MDA8 O_3_SNO_x concentrations in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW.

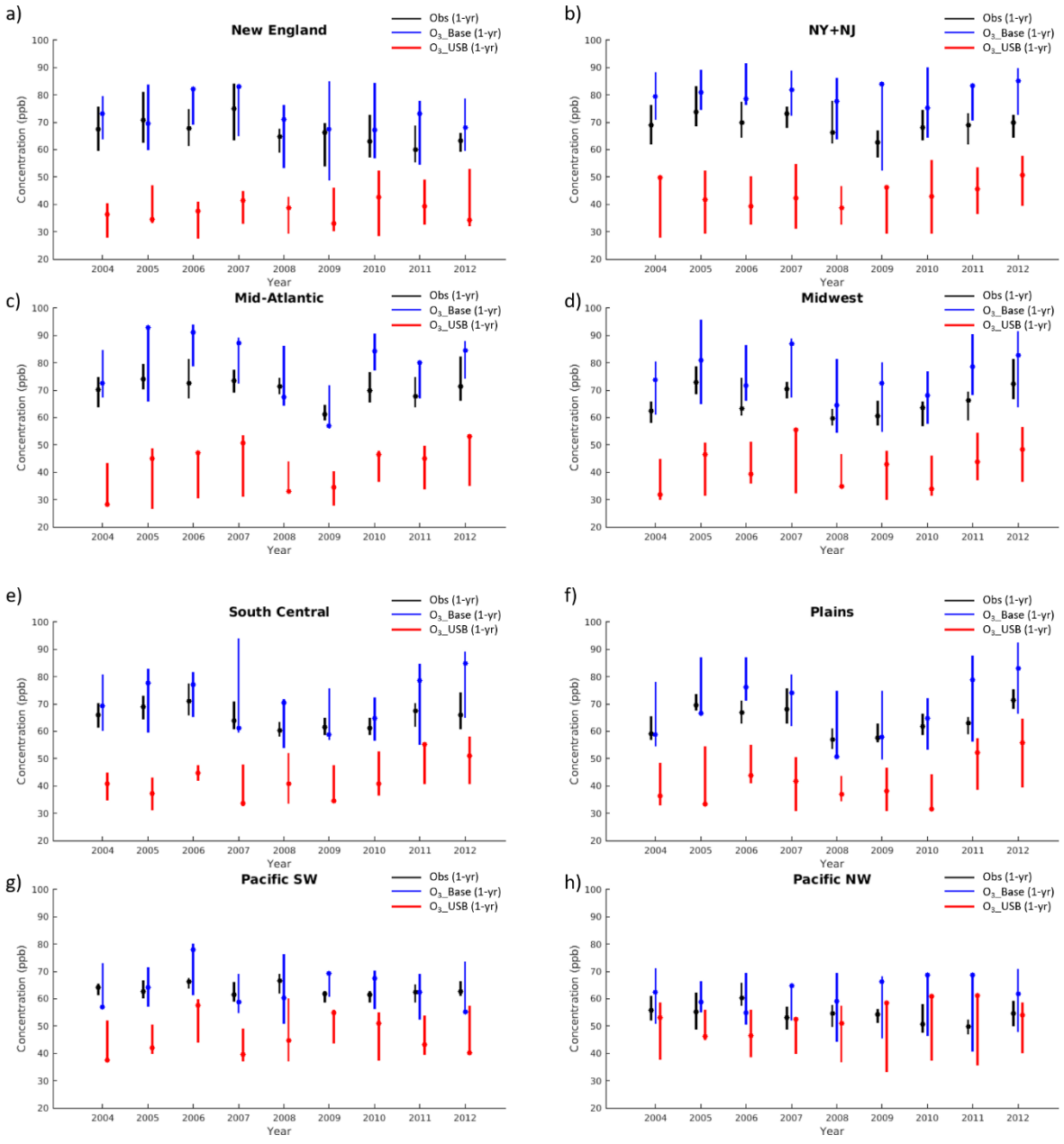


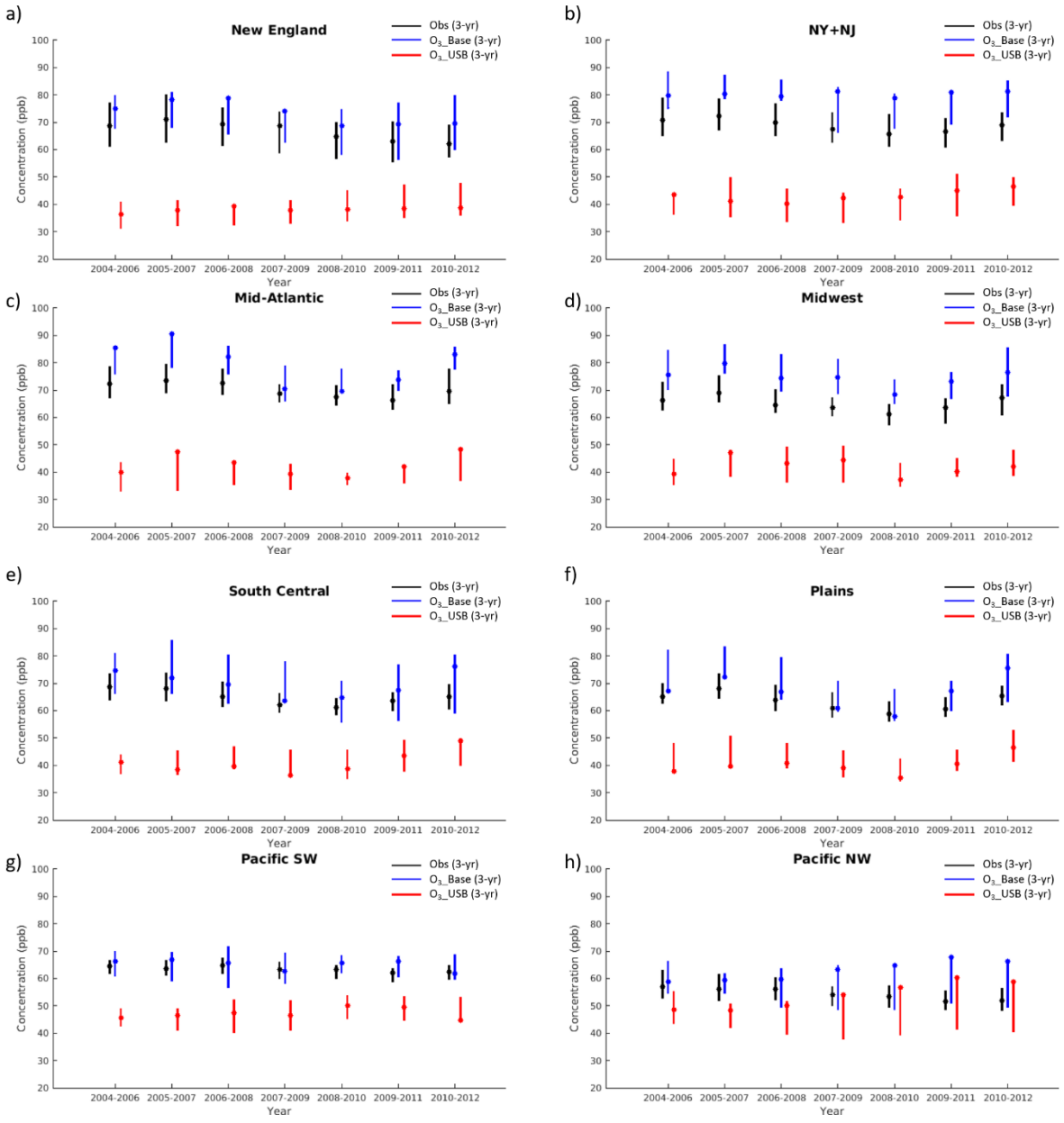
Supplemental Figure 14: Anomaly on the MDA8 O_3_top10obs_JJA days of each sensitivity simulation relative to the 2004-2012 average in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW. Each panel shows the anomaly from observations, O_3_Base, O_3_USB, O_3_USA, and temperature (in degrees C).

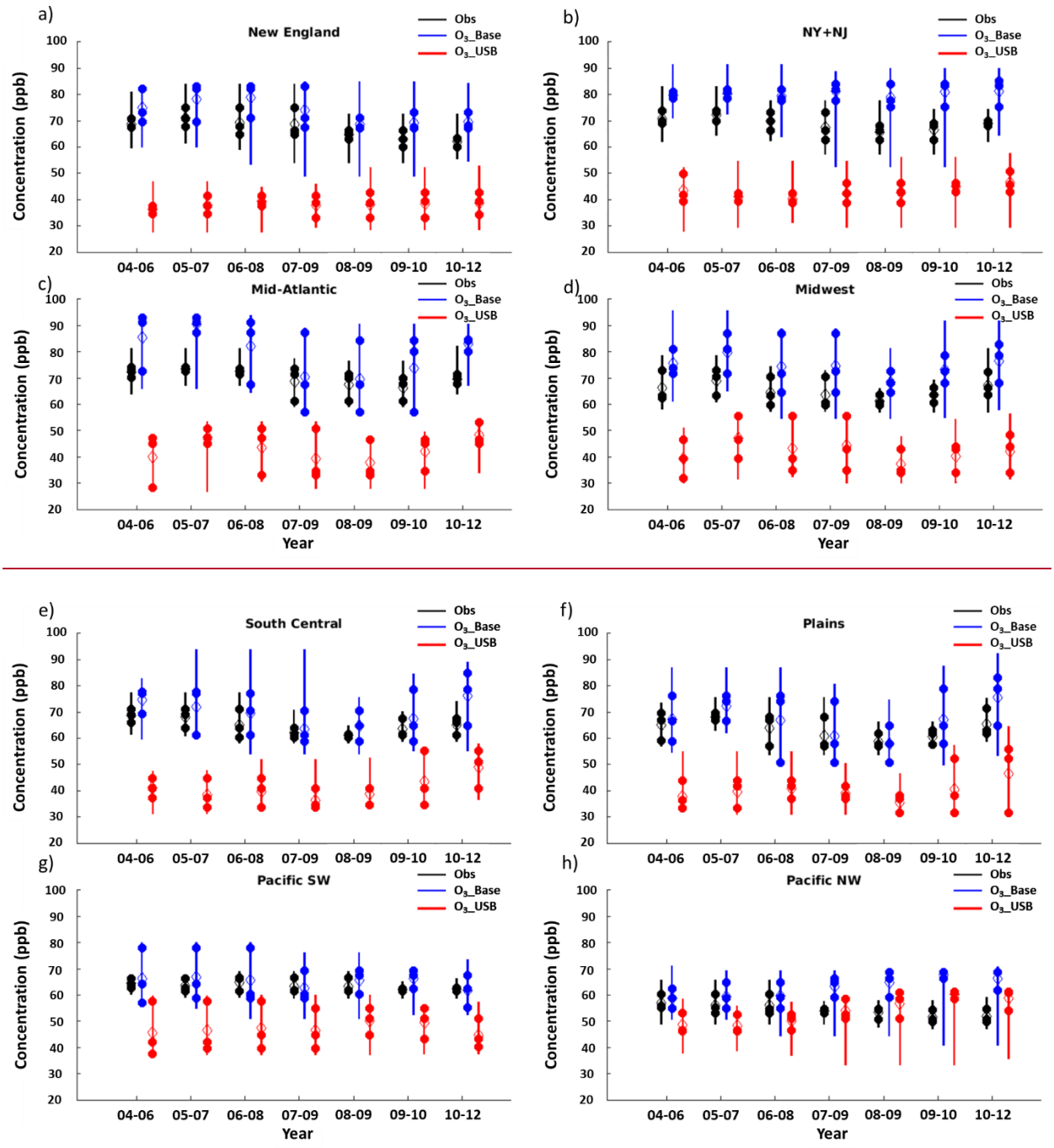


1055

Supplemental Figure 15: Anomaly on the O_3_top10obs_JJA days for each sensitivity simulation relative to the 2004-2012 average in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW. Each panel shows the anomaly from O_3_BVOC, O_3_SNO_x, O_3_$NALNO_x$, O_3_BB, O_3_ICT+CH_4, and O_3_CA+MX.


a)
New England
b)
NY+NJ
c)
Mid-Atlantic
d)
Midwest
e)
South Central
f)
Plains
g)
Pacific SW
h)
Pacific NW
Obs (1-yr)
O3_Base (1-yr)
O3_USB (1-yr)
Concentration (ppb)
Year


a)
New England
b)
NY+NJ
c)
Mid-Atlantic
d)
Midwest
e)
South Central
f)
Plains
g)
Pacific SW
h)
Pacific NW
Obs (3-yr)
O3_Base (3-yr)
O3_USB (3-yr)
Concentration (ppb)
Year
2004-2006
2005-2007
2006-2008
2007-2009
2008-2010
2009-2011
2010-2012



Supplemental Figure 16: Summary information for each region showing the three 4th highest days in each year (solid dots) that went into the calculation of the three-year average of the 4th highest MDA8 O_3 day (hollow diamond). Error bars show the range between the highest and lowest O_3_top10obs days across each 3-year span (i.e, across 30 total points) occurring between March and October in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW. Observations are shown in black, O_3_Base is in blue, and O_3_USB is in red. ~~regions. (c), (d), (e), (f) show the range of the O_3_top10obs days after averaging over three consecutive years. The solid dots show the 4th highest MDA8 O_3 day for each simulation (a, b) and the annual 4th highest MDA8 O_3 day averaged over three consecutive years.: Range in magnitude of the ten highest MDA8 O_3 values for each year shown as vertical lines in the observations (black), O_3_Base (blue), and O_3_USB (red) in (a) New England, (b) NY+NJ, (c) Mid-Atlantic, (d) Midwest, (e) South Central, (f) Plains, (g) Pacific SW, and (h) Pacific NW. a), (b), (e), (f) show the range on of O_3_top10obs days during~~

~~each year between 2004-2012. (c), (d), (e), (f) show the range of the O_3_top10obs days for each year. The solid dots show the 4th highest MDA8 O_3 day for each simulation (a), (b) and the annual 4th highest MDA8 O_3 day.~~

~~**Supplemental Figure 13: Range in magnitude of the ten highest MDA8 O_3 values after averaging over 3 consecutive years in the observations (black), O_3_Base (blue), and O_3_USB (red)**~~

1080