Response to Anonymous Referee #1

The manuscript on US surface ozone trends and extremes by Lin et al. is clearly one of the best modelling studies I have read in my career. It covers an important scientific topic with political relevance and provides an in-depth analysis of US surface ozone and its drivers to the extent that this can be achieved with a global model. It contains a careful and insightful analysis of observations and model results including a well-designed set of sensitivity experiments to attribute ozone trends and variability to various factors. The text is well structured and very well written. All arguments are clearly presented and justified; there is an adequate recognition of previous work. The figures are also very well designed and clear and readable. This would have almost been the first manuscript which I would recommend to "publish as is", except that I do have a few very minor comments and suggestions how the text could be even further improved. In short, it was a real pleasure to review this manuscript.

RE: We truly appreciate the reviewer for carefully reading the manuscript and for favorable comments and insightful suggestions. Below we include a point-by-point response (in bold blue) to the reviewer, responding to *their comments (in italic)* and explaining the changes made to the manuscript.

Introduction: start with at least one general sentence about ozone being an important air pollutant which has been of relevance to the US for a long time.
RE: Good suggestion! We now begin with this sentence:

"Within the United States, ground-level O_3 has been recognized since the 1940s and 1950s as an air pollutant detrimental to public health."

Page 2, lines 7-10: explicitly mention methane here (part of climate effects?)
RE: Done. "There are concerns that rising Asian emissions and global methane …"

Page 2, lines 33/34: this result is based on a previous study with the same model. Don't state it as undisputed fact. Please write "Previous model simulations indicate . . ." or similar.
RE: We now say:

"Model simulations indicate that import of Asian pollution enhances mean WUS surface O_3 in spring by ~5 ppb (Zhang et al., 2008; Lin et al., 2012b), and occasionally contributes 8-15 ppb during springtime pollution episodes observed at rural sites (Lin et al., 2012b) as supported by in situ aerosol composition analysis (VanCuren and Gustin, 2015)"

Page 3, line 2: not only precursor trends, but also inter-annual (meteorological) vari- ability make this difficult if not impossible
RE: Good point! We now say:

"Discerning directly the effect of climate change on air quality from long-term observation records of O_3 would be ideal, but concurrent trends in precursor emissions and large internal variability in regional climate impede such an effort."

Page 3, line 14: you may also want to mention that models have difficulties in simulating the seasonal cycle at baseline sites correctly (see recent papers by Parrish et al., Derwent et al.)
RE: We did not make change here because the focus of this paper is on long-term trends. Adding discussions on the seasonal cycle somewhat interrupts the overall flow of that paragraph.

Section 2: please provide at least one general statement about the GFDL model with a reference to the model description paper before describing the experiments.
Page 4, line 22: please provide a reference to the dry deposition climatology
RE: Done. Please see the revised manuscript.

Page 5, line 22: awkward grammar: "a number of studies (Hiboll)."
RE: Changed to "… by a few recent studies (e.g., *Hilboll et al.*, 2013)"

Page 6, lines 7-10: statement misleading: there are thousands of long-term monitoring sites from AQS and several hundred "rural" stations. Add "selected"?
RE: Add "selected".

Page 6, lines 15-17: Please state if trend was derived from daily MDA8 values or monthly values and how you tested for the appropriateness of a linear trend model.
RE: This is clarified in Section 2.3.

"The trend is calculated separately for the 5th, 50th and 95th percentiles of daily MDA8 O_3 for each season through ordinary linear least-square regression. Statistics are derived for the slope of the linear regression in units of ppb yr-1, the range of the slope with a 95% confidence limit (not adjusted for sample autocorrelation), and the p-value indicating the statistical significance of the trend based on a two-tailed t-test."

Page 9, line 8 vs. Caption Figure 6: Lee et al. once cited as 2013, and once as 2014.
RE: Nice catch! Revised.

Page 10, line 11: Please add a quantitative summary statement how well the Asian trends are captured. Figure 6 indicates within 10-20%, Mt. Happo is within 37%.
RE: Good suggestion! We now stated "We conclude that GFDL-AM3 captures 65-90% of the observed O_3 increases in Asia, lending confidence in its application to assess the global impacts of rising Asian emissions."

Page 11, lines 9-20: I recall from earlier discussions on USNE surface ozone that a large change occurred around 2001 when NOx scrubbers in power plants were activated. Is this worth

mentioning here? Could this have an impact on the observed trends and/or the relation between spring and summer trends?
RE: We now mention this in the revised manuscript:

"Many northeast states in the late 1990s and early 2000s did not turn on power plant NO_x emission controls until the O_3 season (May-September), which may also contribute to observed differences between spring and summer O_3 trends."

Page 13, line 4 vs. 20 ff: perhaps the rising isoprene discussion could be merged in one place? It is slightly confusing to see this in two places.
RE: We have moved that sentence down to the next paragraph.

Page 14, lines 11ff: Figure caption (Figure 13) uses "NAB" as abbreviation for "Background" run. This should be made consistent (also the font of "NAB" in the legend differs from the other legend entries).
RE: The "NAB" abbreviation is only used in Figure 13 because of limited space. We have used the term "Background" throughout the text in the manuscript.

Page 15, line 1: Does the statement "can explain 50-65%..." assume linear additivity of the factors controlling surface ozone? Would the impacts be the same if you applied linear regression on the differences between the model simulations (instead of sub- tracting the linear trend estimates from each other)? Perhaps, Table 2 would be easier to digest if the individual contributions were listed (i.e. the differences) instead of the regression results themselves?
RE: As suggested by the reviewer, we apply linear regression on the differences between the model simulations and find no significant change from the impacts calculated by subtracting the linear trends in Table 2. Thus, no change is made in the manuscript.

Page 15, line 38: please add a note how Asian emissions will decrease after 2030 according to RCP8.5. For example, will they reach year 2000 or year 1990 levels?
RE: Done.

"Under the RCP8.5 scenario, Chinese NO_x emissions are projected to peak in 2020-2030, reflecting an increase of ~50% from 2010 (Fig.1a), followed by a sharp decrease reaching 1990 levels by 2050."

Page 16, lines 33/34: "consistent with the seasonality of pollution transport from Asia." Isn't this also the influence of the Asian summer monsoon in July/August which reduces surface ozone over Asia itself?
RE: We now say:

"The stronger increase measured in June than in July-August is consistent with the Asian summer monsoon producing surface O_3 minimum in July-August (e.g., Lin et al., 2009), as well as the seasonality of intercontinental pollution transport."

Page 20, lines 22-27: if possible, the argument about dry deposition influencing the high end of ozone distributions during the 1988 heatwave should be substantiated by an additional (1-year or only summer months) model simulation where dry deposition could be turned off (or reduced).

RE: Thanks for the suggestion! We have conducted a sensitivity simulation for 1988 with 35% decreases in O_3 deposition velocities over drought areas. The results are shown in Figs. 18 and 19 (copied below) and discussed in Section 6 (please see tracked changes in the revised manuscript).

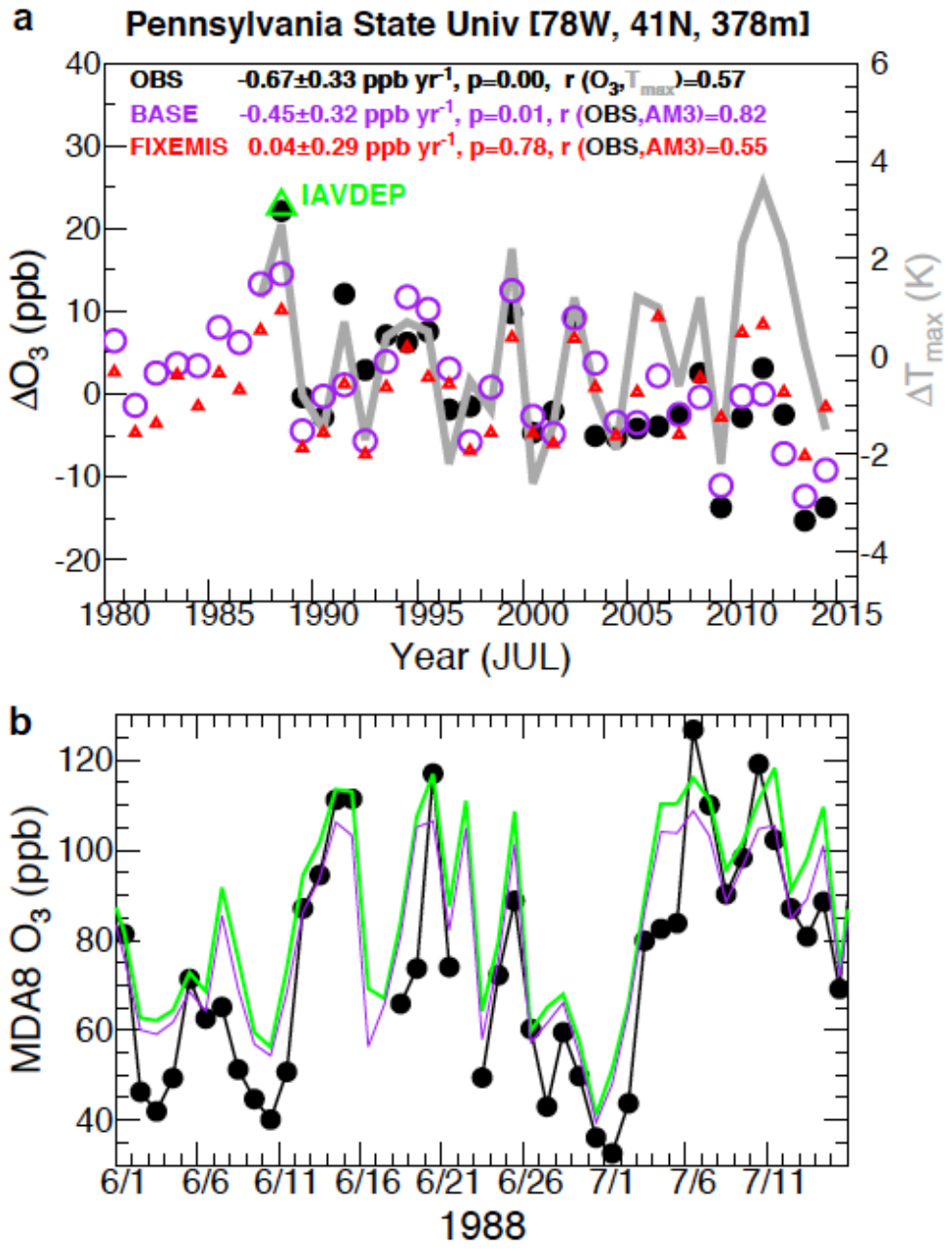

Figure 18. **(a)** Time series of July mean MDA8 O_3 anomalies (relative to 1988-2014) at the Pennsylvania State University (PSU) CASTNET site as observed (black) and simulated by the GFDL-AM3 model with time-varying (purple) and constant anthropogenic emissions (red), along with observed anomalies in July mean daily max temperature (gray lines; right axis). The green triangle denotes the 1988 O_3 anomaly from a sensitivity simulation using BASE emissions but with 35% decreases in $V_{d,O3}$. **(b)** Time series of daily MDA8 O_3 at PSU from June 1 to July 16 in 1988 as observed (black) and simulated by the BASE model (purple) and the sensitivity simulation with 35% decreases in $V_{d,O3}$ (green).

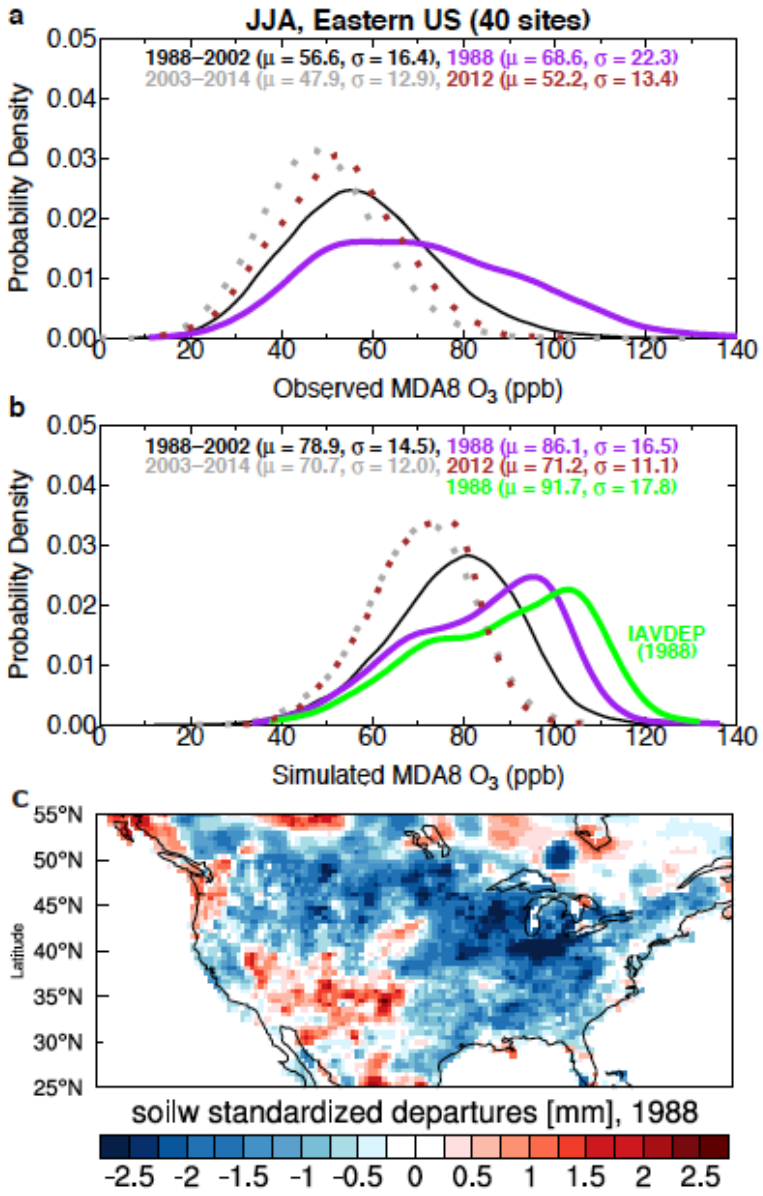

Figure 19. **(a)** Comparisons of probability distributions of summertime MDA8 O_3 from 40 EUS CASTNet sites for the pre-NO_x SIP Call (1988-2002; solid black) versus post-NO_x SIP Call (2003-2014; dashed gray) periods and during the extreme heat waves of 1988 (solid purple) versus 2012 (dashed brown). The median (μ) and standard deviation (σ) are shown (ppb). **(b)** Same as **(a)** but from AM3 BASE. Also shown is the O_3 distribution in 1988 from a sensitivity simulation with 35% decreases in $V_{d,O3}$ in drought areas (green). **(c)** Standardized soil moisture departures for JJA 1988 (calculated by dividing anomalies by the 1979-2010 climatological standard deviation, using data from NOAA Climate Prediction Center).

Page 21, lines 1-2: how about "plume chemistry" as another explanation for the overall bias? There are strong NO_x gradients also in the horizontal, and ozone production efficiency is higher in the medium-NO_x range than in the high NO_x range.

RE: We don't think model limitation in resolving plume chemistry is a major explanation for the bias. Travis et al. (2016) used a 0.25°x0.25° model and found

20 ppb biases similar to our 2-degree model before adjustment of NO_x emissions. No changes are made in the manuscript.

Figure 20: why are the observed trends not included in this figure?
RE: Because this figure shows decadal mean changes from 1981-1990 to 2003-2013. There are only limited observations available during 1981-1990. We have clarified this in the caption of Fig.20.

Conclusions: the conclusions are more a summary than real conclusions. I suggest to shorten this summary of results and instead try to go one step further in assessing the possible consequences of this study. For example: even though methane hasn't played a major role in the past, will it become more important in the future if, as suggested by the RCPs, Asian NOx emissions will decrease again? Or: what do we expect from future NOx emissions in the NEUS? In relation to climate change: could there be a greater role of biogenic VOC and would this lead to more or less severe ozone episodes?

RE: Good suggestion. We have changed the title of Section 7 to "Conclusion and Recommendations". Now we explicitly discuss the implications of our work; (1) on the common model biases on baseline ozone trends and recommendations for future multi-model analysis for the Chemistry-Climate Model Initiative, (2) the growing importance of rising global methane and NOx emissions in South Asian countries, where ozone production is more efficient, after Chinese emissions continue to decline in the coming decades, (3) the benefits of future NO_x emission controls on O_3 reductions in the Southeast US, and (4) uncertainty in model treatment of land-biosphere couplings for projecting pollution extremes in a warming climate. Please see tracked changes in Section 7 of the revised manuscript.