Dear editor and dear reviewers:

On behalf of my co-authors, we thank you very much for giving us an opportunity to revise our manuscript, we appreciate editor and reviewers very much for their positive and constructive comments and suggestions in the ACPD.

We have studied reviewer's comments carefully and have made revision which marked in red in the paper.We have tried our best to revise our manuscript according to the comments. Please see below our replies in detail. Attached please find the revised version, which we hope reviewer would be satisfied with our answers and the revision we provided.

We would like to express our great appreciation to reviewer for comments on our paper. Looking forward to hearing from you.

Yours sincerely,

Shuai Li
College of Environmental Science and Engineering Donghua University
2999 Renmin North Road, Songjiang District, Shanghai
Mobile: (86)13999856917
Email: rainlishuai@163.com

Referee #1

Reviewer comments on Li et al., "Historical and future changes and present-day uncertainties of ozone in China from CMIP6 models"

The paper by Li et al., uses output from multiple CMIP6 models to examine the performance and projection of surface ozone across China. The paper first provides a detailed present-day evaluation of CMIP6 models over China using observational data-fusion product TAP, showing how CMIP6 models tend to underestimate surface ozone over Eastern China but overestimate in the southwest. An analysis is then also performed to look at the influence of cloud, vegetation types and aerosols on the simulation of surface ozone over China. The historical changes in surface CMIP6 models is then discussed showing large increase across China since 1850. Finally, an analysis of future changes in surface ozone across China is presented showing the impact of high mitigation scenarios reduces surface ozone compared to small increases under weak mitigation scenarios. The contains some useful evaluation of present-day surface ozone concentrations across China and also highlights the future changes in different scenarios. I think the manuscript could be published once the comments below have been addressed.

1. Major Comments on sections

Q1. Abstract – At the moment the abstract just reads like a list of results, which is just a slightly shorter version of the conclusion. It would be better if the abstract was set out to following the traditional format of including a bit on the background/introduction, methods, results, and conclusions of the study. The abstract does not currently do this and so is hard to get a good summary of the study presented here.

Response: We thank the reviewer for this helpful suggestion. Following the reviewer's comment, the abstract has been substantially revised to follow a conventional structure, including background, data and methods, key results, and major conclusions, thereby providing a clearer and more balanced summary of the study.

Revised context are shown below (Lines 16–37):

Ozone (O_3) contributes to global climate change and poses a direct threat to human health. This study investigates the historical variability, future projections, and associated present-day uncertainties of surface O_3 concentrations over China using simulations from nine CMIP6 models and observational data from the Tracking Air Pollution in China (TAP) dataset. A multi-model ensemble mean (MME), constructed using an equal-weighted approach, is used to evaluate model uncertainties across different seasons, underlying surface types, total cloud cover, and $PM_{2.5}$ concentrations, and to assess model performance under future Shared Socioeconomic Pathway (SSP) scenarios. The results show that the

MME captures the pronounced seasonal cycle of surface O_3, with higher concentrations during June–August (JJA, ~105 $\mu g \cdot m^{-3}$) and lower values during December–February (DJF, ~55 $\mu g \cdot m^{-3}$), but underestimates O_3 across most regions of China, particularly in East China. Model performance varies with environmental conditions, showing lower bias, MAE, and RMSE over natural land surfaces than over anthropogenic surfaces. The O_3 bias is minimized under cloudy conditions, maximized under partly cloudy conditions, and generally increases with rising $PM_{2.5}$ concentrations before declining beyond a threshold. Over the historical period, the MME simulates a substantial increase in annual mean surface O_3 across China (~39.3 $\mu g \cdot m^{-3}$). Future projections indicate continued O_3 increases under weak mitigation (SSP3-7.0), with East China rising by 26.9%, and widespread decreases under strong mitigation (SSP1-2.6), particularly in Southwest and South China (>30 $\mu g \cdot m^{-3}$). Analysis of model spread and its drivers indicates that uncertainties in surface O_3 projections arise from the combined effects of emissions (including precursors and $PM_{2.5}$), climate conditions, and model representations of chemistry and circulation. Improving the understanding of these coupled influences is essential for enhancing the reliability of regional O_3 projections and for informing effective air quality and climate mitigation strategies in China.

Q2. Methods – I think more detail should be included in the methods to help the reader understand how the study was performed.

(1) Can you provide a list of which CMIP6 variables that are used in the study and any data references, perhaps in a supplementary table.

Response: Thank you for this helpful suggestion. We agree that clear documentation of the variables and data sources is essential for transparency and reproducibility. In response, we have added a new supplementary table (Table S1. CMIP6 variables and auxiliary datasets used in this study). Although the reviewer specifically requested a list of CMIP6 variables, Table S1 provides a comprehensive overview of both CMIP6 model outputs and auxiliary datasets employed throughout the study. The table lists the variable names, data sources, temporal resolution, purpose in the analysis, and corresponding references. Revised Tables are shown below:

Table S1. CMIP6 variables and auxiliary datasets used in this study.

Variable name	Data source	Temporal resolution	Purpose in the analysis	Reference
Ozone	CMIP6 models	Amon	O_3 analysis	Eyring et al. (2016)
	TAP dataset	Mon	Uncertainties analysis	Xue et al. (2020)
	Ground-based observations	Mon (derived from hourly)	O_3-NO_2-NMVOCs relationship	Li et al. (2017)
NO_x	CMIP6 models	Amon	Correlation analysis	Collins et al. (2017)
	Ground-based observations	Mon (derived from hourly)	O_3-NO_2-NMVOCs relationship	Li et al. (2017)
TAS	CMIP6 models	Amon	Correlation analysis	Eyring et al. (2016)
CH_4	CMIP6 models	Amon	Correlation analysis	Collins et al. (2017)
NMVOCs (emivoc)	CMIP6 models	Amon	Correlation analysis	Collins et al. (2017)
BVOCs (emibvoc)	CMIP6 models	Amon	Correlation analysis	Collins et al. (2017)
Surface types	MODIS/MCD12Q1	2020 (fixed)	Uncertainties analysis	Friedl and Sulla-Menashe (2019)
Total cloud cover	ERA5 (ECMWF)	Mon	Uncertainties analysis	Wu et al. (2023)
$PM_{2.5}$ and its components	TAP dataset	Mon	Uncertainties analysis	Xiao et al. (2022)

(2) Is it worth including any reference to data from the Tier2 scenarios due to the limited available data and also the problem of comparing scenarios when the models providing results are not consistent? Unless there is one model that provides data for all and there is a useful comparison here?

Response: We thank the reviewer for this insightful comment. We fully agree that the limited availability and inconsistent model coverage of Tier 2 CMIP6 scenarios make robust quantitative inter-scenario comparisons challenging. In this study, all core quantitative analyses—including present-day uncertainty assessment, historical evaluation, and multi-model spread analysis—are conducted exclusively using Tier 1 scenarios, which provide the most extensive and consistent multi-model coverage. Tier 2 scenarios (SSP1-1.9, SSP4-3.4, SSP4-6.0, and SSP5-3.4-OS) are included in Section 5.1 only to illustrate potential trajectories of future surface O_3 under alternative emission pathways. Although one CMIP6 model provides simulations for all Tier 1 and Tier 2 scenarios, single-model inter-scenario comparisons are not performed, because the study relies on multi-model ensemble statistics to ensure representative and robust conclusions.

To avoid any ambiguity, the Research Data and Methods section has been revised to clarify the distinct roles of Tier 1 and Tier 2 scenarios (Lines 165–174 in the revised manuscript).

"Due to the limited availability and inconsistent model coverage of Tier 2 CMIP6 scenarios (SSP1-1.9, SSP4-3.4, SSP4-6.0, and SSP5-3.4-OS), all present-day uncertainty analyses and multi-model spread analysis in this study are conducted exclusively using Tier 1 scenarios. Tier 2 scenarios are included in Section 5.1 only to illustrate potential future trajectories of surface O_3 under alternative emission pathways. Some numerical changes are reported to highlight relative differences between these scenarios, but they are not used for robust multi-model inter-scenario comparisons or quantitative uncertainty assessment. Although one CMIP6 model (MRI-ESM2-0) provides simulations for all Tier 1 and Tier 2 scenarios, single-model inter-scenario comparisons are not performed, as our analysis relies on multi-model ensemble statistics to ensure representative and robust conclusions."

(3) Could the results from TAP be included in a first section of the results along with a direct comparison to the CMIP6 results?

Response: We thank the reviewer for this thoughtful suggestion. In our study, TAP is primarily used as an independent reference to evaluate CMIP6 multi-model simulations under different environmental conditions, rather than as a parallel simulation framework.

Accordingly, the Results section is structured such that Section 3.1 focuses on the inter-model performance and spatiotemporal evaluation of present-day CMIP6 simulations, while the direct comparisons between TAP and CMIP6 are systematically presented in Sections 3.2-3.4 under different underlying surface types, cloud cover levels, and $PM_{2.5}$ concentrations and components.

To further clarify this structure, we have revised the title of Section 3.1 and added explanatory text at the beginning of Section 3 describing the role of TAP in the comparative analysis.

Please see Lines 226–229 in the revised manuscript:

"In this section, we first evaluate the inter-model performance and spatiotemporal characteristics of CMIP6 surface O_3 simulations. TAP is then used as an independent reference to assess CMIP6 uncertainties under different environmental conditions, including underlying surface types, cloud cover levels, and $PM_{2.5}$ concentrations and components (Sections 3.2–3.4)."

(4) An ensemble mean for each model is first performed before using to compare to other models. What is variability in surface ozone across ensemble members for a particular model like?

Response: We thank the reviewer for this valuable comment. To quantify internal variability, we

estimated the inter-ensemble variability of monthly mean surface O_3 for CMIP6 models with more than two ensemble members under SSP3-7.0. For each model, the standard deviation of monthly mean O_3 across ensemble members was computed for each calendar month and then averaged over 2015–2023 (Table S6).

The resulting standard deviations are generally 2–4 $\mu g \cdot m^{-3}$, with a multi-model mean of ~2.9 $\mu g \cdot m^{-3}$. This magnitude is small compared with the inter-scenario differences discussed in this study, indicating that it does not compromise the robustness of the multi-model comparison based on ensemble means.

Added tables in supplement are shown below:

Table S6. Inter-ensemble variability of surface O_3 under SSP3-7.0 (2015–2023). Quantified as the average monthly SD across ensemble members for each model ($\mu g \cdot m^{-3}$).

Models	BCC-ESM1	CESM2-WACCM	EC-Earth3-AerChem	MRI-ESM2-0	UKESM1-0-LL	MME
SD (O_3)	3.3	2.4	2.6	3.9	2.4	2.9

Note: Only models with more than two ensemble members under SSP3-7.0 were included.

Please see Lines 205−210 in the revised manuscript:

"Internal variability due to ensemble spread was quantified for CMIP6 models with more than two realizations under SSP3-7.0. For each model, the standard deviation (SD) of monthly mean surface O_3 across ensemble members was calculated for each month and then averaged over 2015–2023 to yield a single metric (Table S6). Ensemble means of each model were used for inter-model comparisons. The internal variability (generally 2–4 $\mu g \cdot m^{-3}$) is smaller than inter-model differences, indicating that it does not compromise the robustness of the multi-model comparison based on ensemble means."

(5) The CMIP6 model output (>100 km) is linearly interpolated to that of the TAP data (0.1°). How does the linear interpolation of coarse model data impact on the concentrations produced? Also is there a limitation of using global models at >100km resolution to represent pollutant conditions over highly urbanised regions?

Response: We thank the reviewer for this important comment.

Regarding the first point, linear interpolation of coarse-resolution CMIP6 model outputs (>100 km) to a 0.1° grid does not introduce physically meaningful fine-scale spatial variability. Instead, it primarily results in a spatial smoothing of the concentration fields. This interpolation preserves the large-scale spatial patterns simulated by the global models, but inevitably smooths localized gradients, leading to

an underestimation of peak concentrations, particularly over highly urbanized regions.

Regarding the second point, we acknowledge that global chemistry-climate models with horizontal resolutions coarser than ~100 km have inherent limitations in representing air pollutant concentrations over highly urbanized areas, as they cannot explicitly resolve urban-scale emission heterogeneity and associated fine-scale processes. As a result, simulated concentrations tend to be more representative of regional- to continental-scale conditions rather than city-scale variability.

We acknowledge this limitation and have added a statement in the Methods section clarifying that the interpolated concentrations are therefore more representative of regional-scale patterns rather than urban-scale variability. In addition, this limitation is further discussed in the Discussion section, where we emphasize that the results of this study are most suitable for analyzing regional-scale trends and multi-model mean behavior, rather than city-scale exposure assessments.

Please see Lines 198–204 and Lines 990–996 in the revised manuscript:

"It should be noted that CMIP6 outputs, originally at coarse spatial resolutions (>100 km), retain only large-scale spatial patterns through linear interpolation and do not introduce physically meaningful fine-scale features. As a result, localized concentration peaks, particularly in highly urbanized regions, may be smoothed or underestimated, and the interpolated concentrations are therefore more representative of regional-scale patterns rather than urban-scale variability. This approach has been widely adopted in previous CMIP6-based air quality studies when comparing global model outputs with higher-resolution datasets (Turnock et al., 2020; Su et al., 2022)."

"This study is based on the CMIP6 multi-model ensemble mean. The MME approach integrates information from different models and mitigates the influence of individual model biases, yet it carries certain limitations. First, the CMIP6 global models operate at coarse spatial resolutions (>100 km), which limits their ability to resolve fine-scale urban features, emission heterogeneity, and local pollution hotspots. Consequently, peak pollutant concentrations in highly urbanized regions may be underestimated, and the results are therefore more suitable for analyzing regional-scale trends rather than city-scale exposure assessments."

(6) How is the present day uncertainty analysis performed? This seems quite vague and where have the information on temperature, clouds and land surface been obtained from? Is this from the models themselves (which would add additional uncertainty) or somewhere else? Also how does the interpolation of global model data impact this comparison on land surface types?

Response: We thank the reviewer for this helpful comment, which prompted us to clarify the methodology in the revised manuscript. Our responses to the specific points are as follows:

- Methodology and Data Sources:

The present-day uncertainty analysis compares CMIP6-simulated surface O_3 with TAP observations under different environmental conditions (land surface type, cloud cover, $PM_{2.5}$ levels). To avoid introducing additional uncertainty from CMIP6 model variables, we use independent datasets: total cloud cover from ERA5 reanalysis, land surface types from MODIS/MCD12Q1 (treated as fixed categories), and $PM_{2.5}$ data from TAP (Table S1).

- Impact of Interpolation on Land Surface Comparison:

Continuous CMIP6 O_3 fields are linearly interpolated to the 0.1° TAP grid. Categorical land surface types are resampled to the target grid using a nearest-neighbor method, which preserves their original classification without introducing spurious mixed categories. Thus, interpolation affects only continuous O_3 fields (potentially smoothing fine-scale gradients) while leaving the land surface categories used for stratification unchanged. This approach preserves the integrity of the grouping analysis and ensures a robust comparison across different land types.

- Quantification of Uncertainty:

For each group (defined by land type, cloud cover, or $PM_{2.5}$ level), the difference between CMIP6 MME and TAP observations is calculated at each grid cell. The bias, MAE, and RMSE of these differences serve as our uncertainty metrics.

This clarification has been added to the Methods section in Lines 211−224 as follows.

"The present-day uncertainty of CMIP6-simulated surface O_3 concentrations is assessed by comparison with TAP observations under different environmental conditions, including underlying surface types, total cloud cover, and $PM_{2.5}$ concentrations and its components. To avoid introducing additional uncertainty from the CMIP6 models themselves, independent datasets are used for stratification: Total cloud cover from the ERA5 reanalysis, underlying surface types from the MODIS/MCD12Q1 product (treated as fixed categories), and $PM_{2.5}$ data from the TAP observations (Table S1). To ensure spatial consistency, continuous CMIP6 O_3 fields are linearly interpolated onto the 0.1° TAP grid, while categorical underlying surface types are resampled to the same grid using a nearest-neighbor method to preserve their classification. Therefore, interpolation affects only continuous model fields and does not alter land surface categories, although it may smooth fine-scale gradients in modeled concentrations. To quantify uncertainties, for different underlying surface type, cloud cover category, and $PM_{2.5}$ concentration and components, the difference between CMIP6 MME and TAP observations is calculated at each grid cell. The bias, MAE, and RMSE of these differences within each group are used as metrics of model uncertainty."

(7) Also you mention in section 3.1 that the MME for CMIP6 models is calculated over the period 2014-2023, which CMIP6 experiments has data been obtained from as this period spans both the historical (up to 2015) and scenario (from 2015 to 2100) experiments. More information needed on this.

Response: Thank you for the reviewer's comment. We realize that there was an error in the original manuscript. The multi-model ensemble mean (MME) in Section 3.1 was calculated over the period 2015−2023, rather than 2014−2023 as previously stated. All CMIP6 data used for this period were obtained exclusively from the scenario experiments, as 2015 marks the beginning of the SSP simulations in CMIP6. No historical experiment data were used in the calculation of the MME for Section 3.1. For each SSP scenario (e.g., SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5), the MME was constructed independently using the available models providing data for that scenario.

We have corrected the time period and clarified the corresponding CMIP6 experiments in the revised manuscript. We apologize for the initial error and thank the reviewer for their careful reading, which has improved the clarity and accuracy of the manuscript.

Q3. Results –

(1) Including a table of statistics to include bias, correlations, trends etc would help improve visibility of results for both the present-day evaluation

Response: We thank the reviewer for the suggestion. In response, we have added the relevant statistics in the Supplementary Materials: Table S2 summarizes the seasonal performance statistics of the multi-model ensemble (MME) O_3 relative to TAP, and Table S3 provides the annual correlation coefficients (r) between individual CMIP6 models and TAP.

Revised Tables are shown below:

Table S2. Summary of seasonal performance statistics of MME O_3 relative to TAP

Statistics ($\mu g \cdot m^{-3}$)	MAM	JJA	SON	DJF
Bias	-25.4	-12.8	-22.9	-25.3
SD (Inter-model)	16.8	18.8	15.2	19.0

Table S3. Annual correlation coefficients (r) between individual CMIP6 models and TAP.

Models	BCC-ESM1	CESM2-WACCM	EC-Earth3-AerChem	GFDL-ESM4	IPSL-CM5A2-INCA	MIROC-ES2H	MRI-ESM2-0	UKESM1-0-LL	UKESM1-1-LL	MME
r	0.78	0.91	0.77	0.80	0.80	0.88	0.81	0.73	0.74	0.80

(2) There is a lot of detail on the land surface section about each land cover type, which the global models might have large uncertainties with their process representation. Therefore, I am not sure I get the point of the land cover section other than to highlight models are different from observations due to emissions/deposition processes, especially with the uncertainties in the ability of models to represent these proceses. Is there anything specific to East Asia?

Response: Thank you for the comment. We acknowledge that global models have large uncertainties in representing detailed land surface processes. The purpose of this section is not to claim precise quantitative representation, but to provide a diagnostic perspective on how different land cover types may influence surface O_3 simulations (e.g., cropland, urban, snow/ice surfaces). The added discussion clarifies the meaning of analyzing land cover impacts in understanding O_3 simulation uncertainties and potential improvements in model parameterizations.

Please see Lines 437–448 in the revised manuscript:

"The simulation biases identified over different surfaces highlight potential challenges for modeling surface O_3 in China. Specifically, the high errors over cropland and urban areas likely reflect the need for improved model representation of regionally characteristic anthropogenic activities and heterogeneous land cover. Meanwhile, uncertainties over snow and ice surfaces point to gaps in modeling cold-season processes, critical for high-altitude and northern parts of China. While global models may not resolve all local processes, analyzing biases by land cover type provides crucial insight into how these regionally characteristic processes, such as the interplay of anthropogenic and biogenic emissions, deposition, and meteorology, govern surface O_3 patterns. This regional perspective helps interpret model-observation discrepancies in China and provides context for understanding O_3 simulation uncertainties related to land cover. These findings suggest that enhancing parameterizations for characteristic land-surface processes could improve the reliability of global and regional models applied."

(3) What is the source of the cloud cover data and are there uncertainties in this? Can you separate out any analysis into clear sky and cloudy sky?

Response: Thank you for the comment. In the revised manuscript, the total cloud cover used for classifying clear-sky and cloudy-sky conditions has been clarified. Total cloud cover was obtained from the ERA5 reanalysis ,and it provides physically consistent cloud fields constrained by observations, and is therefore more suitable for diagnosing the observationally based cloud effects on surface O_3 biases. The cloud-cover classification (clear sky, partly cloudy, cloudy, overcast) follows Han and Cong (2015). We also note that uncertainties exist in reanalysis cloud cover due to retrieval errors and model assumptions.

This point has been added to the revised manuscript in Lines 455–463 as follows:

"The analysis of surface O_3 bias between MME and TAP under different cloud conditions (Fig. 8) is based on total cloud cover from the ERA5 monthly reanalysis dataset. Using these data, and following the classification of Han and Cong (2015), conditions were categorized into four levels: clear sky (0–10%), partly cloudy (20–30%), cloudy (40–70%), and overcast (80–100%). It is important to note that while ERA5 provides observation-constrained, physically consistent fields suitable for this diagnostic purpose, uncertainties inherent to any reanalysis (e.g., from retrieval errors and model assumptions) remain. Nonetheless, ERA5 is considered one of the most reliable datasets for studying cloud-meteorology interactions (Wu et al., 2023), supporting the robustness of our comparative analysis."

(4) Section 3.4 it might be good to mention the main mechanisms that aerosols can impact ozone at the start of the section

Response: Thank you for the comment. We have revised the text to specify the main mechanisms that aerosols affect O_3 formation. The changes can be found in Lines 493–499 as follows:

"Aerosols play a crucial role in the simulation of surface O_3 concentrations. Variations in $PM_{2.5}$ can influence O_3 formation through several pathways: (1) chemical effects: changes in aerosol composition (e.g., sulfate, nitrate, organic matter, BC) can perturb the concentrations of key radicals (e.g., OH, HO_2) and thus alter photochemical reaction rates (Wang et al., 2019); (2) radiative effects: aerosols modify optical properties and reduce actinic flux at the surface, thereby suppressing photolysis rates and slowing O_3 production; and (3) heterogeneous effects: reactions on aerosol surfaces can directly consume NO_x or other precursors, shifting the chemical environment for O_3 formation (Qu et al., 2021)."

(5) Section 4 – Are the discrepancies in the historical trends driven by the 1850 values in each model, as there is larger uncertainty here and also a larger spread in model sensitivities?

Response: Thank you for the comment. The discrepancies in historical O_3 trends among CMIP6 models are influenced by both the initial pre-industrial O_3 values around 1850 and differences in model sensitivities to precursor emissions and aerosol processes. We have clarified this point in the revised manuscript Lines 590–610 as follows:

"Central China exhibits the largest discrepancies in historical surface O_3 changes among the CMIP6 models, with the maximum inter-model difference reaching 29.6 $\mu g \cdot m^{-3}$. This pronounced inter-model spread can be understood in light of previous findings. As highlighted by Turnock et al. (2020), such

variability in CMIP6 surface O_3 responses primarily arises from two factors: (i) differences in the simulated 1850 baseline O_3 concentrations, and (ii) differences in regional O_3 trend rates, reflecting model-dependent chemical sensitivities to precursor emissions, particularly the nonlinear O_3-NO_x relationship. These mechanisms provide a consistent explanation for why models such as MIROC-ES2H simulate much stronger historical increases in surface O_3 than models like UKESM1-0-LL. Importantly, inter-model differences in chemical sensitivity extend beyond the response to NO_x alone, encompassing the treatment of other precursors and, critically, aerosol-chemistry interactions. In Central China, precisely where the inter-model spread is greatest, large historical changes in $PM_{2.5}$ concentrations (Su et al., 2022) may further modulate O_3 formation by influencing heterogeneous radical loss on aerosol surfaces. This aerosol modulation effect is likely most pronounced during seasons characterized by high $PM_{2.5}$ loadings and weak photochemical activity (e.g., winter), when reduced solar radiation and enhanced atmospheric stability may amplify the impact of aerosol-related processes on surface O_3.

In summary, the substantial inter-model spread in historical surface O_3 trends over China since 1850 reflects a combination of uncertainties associated with the pre-industrial baseline state and model-dependent chemical sensitivities. These sensitivities extend beyond precursor emissions alone, encompassing differences in chemical mechanisms and aerosol-chemistry interactions, which can further modulate regional O_3 responses. This highlights the need for future multi-model attribution and constraint-based approaches to explicitly disentangle baseline-related uncertainty from process-level differences."

(6) Section 5 – Could you compare some of your changes in the scenarios to any other similar studies, perhaps over wider regions? Be careful with the description of the SSP scenarios and what they incorporate. Is it better to just focus on results from the main scenarios with the most data e.g. SSP 126, 245, 370, lowNTCF and 585 due to the data availability in the other scenarios? Can you better link the explanations in the model variabilities to the correlation plot on Figure 14?

Response: Thank you for the constructive suggestions.

- **Comparison with other studies:**

We have substantially revised the Discussion section by adding a new subsection and a supplementary table (Table S5 in the Supplement), which compare the projected changes under different SSP scenarios with those reported in previous CMIP6-based studies at broader spatial scales (e.g., East Asia and the global scale). This comparison demonstrates that our main findings are generally consistent

with large-scale assessments, while also highlighting distinct regional characteristics over China. These additions have been incorporated into Section 5.1 (Lines 968–989) as follows:

Overall, the results of this study are broadly consistent with conclusions from CMIP6-based global-scale analyses (Turnock et al., 2020; Griffiths et al., 2021; Zeng et al., 2022), while further revealing more pronounced regional heterogeneity. Consistent with global studies, surface ozone concentrations over China decrease markedly under the strong mitigation scenario (SSP1-2.6), but exhibit overall or episodic increases under high-emission scenarios (SSP3-7.0 and SSP5-8.5). However, under the SSP3-7.0-lowNTCF scenario, which emphasizes reductions in short-lived climate forcers, surface O_3 over China still shows an increasing tendency, in clear contrast to the global mean response. This discrepancy highlights the important role of regional chemical processes and emission structures in modulating ozone changes. Furthermore, this study finds that the magnitude of ozone changes over China is generally larger than the global average, accompanied by more pronounced spatial heterogeneity (Table S5). In terms of chemical mechanisms, the significant negative correlation between O_3 and NO_x over China—indicating the dominance of NO_x titration—is consistent with established understanding for East Asia, whereas global-scale analyses suggest substantial regional heterogeneity in this relationship (Turnock et al., 2020). The positive correlations between surface O_3, CH_4, and TAS are also consistent with CMIP6-based global studies. In contrast, the relationships between O_3 and NMVOCs and BVOCs exhibit pronounced complexity and substantial inter-model variability, which is consistent with findings from global multi-model analyses identifying O_3-VOC sensitivity as one of the key sources of uncertainty in future ozone projections. Consequently, this study not only confirms the applicability of the global scenario framework to the China region, but also quantitatively characterizes the magnitude and uncertainty of ozone changes within key high-emission regions, underscoring the importance of regionally refined process-based analyses in future air quality assessments.

Revised Tables are shown below:

Table S5. Multi-model mean changes in surface ozone concentrations over China and globally under historical and SSP scenarios .

Main Scenario	China (this study; μg·m^{-3})	Global (Turnock et al., 2020; ppb)
Historical	Increase of 39.3±14.4 since 1850	Increase of 11.7±2.3 since 1850
SSP1-2.6	Decrease of 12.6±3.1 by 2050 Decrease of 25.3±7.2 by 2100	Decrease of 5±1.2 by 2050 Decrease of 9±1.6 by 2100
SSP2-4.5	Decrease of 13.6±7.2 by 2100	Decrease of 4±1.7 by 2100
SSP3-7.0	Increase of 8.4±2.0 by 2050 Increase of 3.9±4.0 by 2100	Increase of 1.6±0.9 by 2050 Increase of 0.6±1.0 by 2100
SSP3-7.0-lowNTCF	Increase of 5.8±1.5 by 2050 Increase of 4.9±2.0 by 2100	Decrease of 2.5±0.5 by 2050
SSP5-8.5	Increase of 6.3±1.6 by 2050 Decrease of 3.4±2.9 by 2100	Increase of 1.4±0.8 by 2050 Decrease of 2.7±1.5 by 2100

Note: Values represent the multi-model mean ± one SD. Historical changes are reported relative to pre-industrial conditions (since 1850). Future scenario changes are referenced to 2015–2023 for this study and to 2005–2014 for the global estimates from Turnock et al. (2020). Units differ between regional (μg·m^{-3}) and global (ppb) estimates; comparisons are intended to highlight relative magnitude and directional consistency rather than absolute values.

- **Clarification of SSP scenario descriptions:**

We have revised the text in Section 5.1 to more clearly describe the key features of each SSP scenario and have added Table S4 to the supplementary material for further clarification.

We fully agree that the focus for quantitative analysis should be on scenarios with the most data. Therefore, all core analyses in this study are based exclusively on the Tier 1 scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5, and SSP3-7.0-lowNTCF), which have the highest model consistency. The Tier 2 scenarios are included only in a limited, qualitative context in Section 5.1 to illustrate a broader range of potential futures, and are not used for quantitative inter-scenario comparisons. This distinction is now explicitly stated in the Research Data and Methods section (Lines 165–174) as follows:

"Due to the limited availability and inconsistent model coverage of Tier 2 CMIP6 scenarios (SSP1-1.9, SSP4-3.4, SSP4-6.0, and SSP5-3.4-OS), all present-day uncertainty analyses and multi-model spread analysis in this study are conducted exclusively using Tier 1 scenarios. Tier 2 scenarios are included in Section 5.1 only to illustrate potential future trajectories of surface O_3 under alternative emission

pathways. Some numerical changes are reported to highlight relative differences between these scenarios, but they are not used for robust multi-model inter-scenario comparisons or quantitative uncertainty assessment. Although one CMIP6 model (MRI-ESM2-0) provides simulations for all Tier 1 and Tier 2 scenarios, single-model inter-scenario comparisons are not performed, as our analysis relies on multi-model ensemble statistics to ensure representative and robust conclusions."

Add tables are shown below:

Table S4. Meaning and characteristics of typical scenarios in CMIP6 in this study

（Global warming levels are based on the IPCC AR6 assessments of CMIP6 model projections (2081–2100 vs. 1850–1900) with likely (5–95 %) ranges. Scenarios marked * are research variants and not part of the five core SSP scenarios.）

Scenario	Socioeconomic Pathway	2100 Radiative Forcing ($W \cdot m^{-2}$)	Global Warming in 2100 (°C, relative to 1850–1900)	Emission Trends & Policy Features	Typical Applications / Notes	Reference
SSP1-1.9	Sustainable development (green growth, low inequality, high tech)	≈1.9	~1.5°C (likely range: 1.2–1.7°C)	Strong mitigation; net-zero before 2050	1.5°C pathway for Paris Agreement evaluation	Riahi et al. (2017)
SSP1-2.6	Sustainable development (same as SSP1)	≈2.6	~1.8°C (likely range: 1.37–2.4°C)	Clean energy transition; moderate mitigation	Low-emission sustainable development scenario	IPCC AR6 WGI (2021)
SSP2-4.5	Medium pathway (current trends, moderate mitigation)	≈4.5	~2.7°C (likely range: 2.1–3.5°C)	Limited global cooperation; gradual emission slowdown	Baseline "medium emission" scenario	Riahi et al. (2017)
SSP3-7.0	Regional rivalry (fragmented development, low cooperation)	≈7.0	~3.6°C (likely range: 2.8–4.6°C)	Developing countries grow rapidly; fossil fuels dominate	Used as pessimistic reference scenario	IPCC AR6 WGI (2021)
SSP3-7.0-lowNTCF*	Same as SSP3-7.0 but with NTCF control	≈7.0	Not officially assessed in AR6	Stronger control of CH_4, NMVOCs, BC	Research variant; not an IPCC core scenario; temperature response varies by model	Lund et al. (2020)
SSP4-3.4	Inequality pathway (widening North–South gap)	≈3.4	~2.3°C (likely range: 1.9–2.8°C)	Developed countries reduce emissions; developing nations increase	Medium-low emission; studies of inequality impacts	O'Neill et al. (2017)
SSP4-6.0	Same as SSP4 but with higher global emissions	≈6.0	~3.0°C (likely range: 2.8–4.0°C)	Strong disparity between high- and low-emission nations	Medium-high emission; reflects global inequality	O'Neill et al. (2017)
SSP5-8.5	Fossil-fueled growth (high growth, no mitigation)	≈8.5	~4.4°C (likely range: 3.8–5.7°C)	Very high emissions; no climate policies	Extreme high-emission "worst case" scenario	Riahi et al. (2017)
SSP5-3.4-OS*	Fossil-fueled development (tech-optimistic); overshoot pathway	≈3.4 (overshoot)	Not officially assessed in AR6	High early emissions; large-scale carbon removal later	Overshoot research scenario; peak warming >2°C then declines; not AR6 official estimate	O'Neill et al. (2016)

- **Linking model variability to Figure 14:**

We strengthened the connection between the discussion of inter-model differences and the correlation analysis shown in Figure 14. Specifically, we added text in Section 5.2 explaining how the correlations between O_3 and drivers (e.g., CH_4, TAS, NO_x, NMVOCs, BVOCs) help explain model-to-model variability under SSP3-7.0. This new discussion can be found in Lines 872–905 as follows:

"Figure 14 reveals that under the SSP3-7.0 scenario, the correlations between surface O_3 and its driving

variables show both consistent patterns across CMIP6 models and region-dependent variability. The robust negative O_3-NO_x correlation across all models quantitatively confirms the dominance of NO_x titration in China over the study period (2015–2100). However, the magnitude of this correlation varies substantially ($r = -0.92$ in BCC-ESM1 over Central China and -0.11 in MIROC-ES2H over South China). This spectrum of correlation strengths is a key diagnostic of inter-model differences in representing O_3 photochemistry. Specifically, the stronger negative correlations indicate that these models simulate a more extensive or persistent VOC-limited chemical regime under the SSP3-7.0 pathway. Consequently, the projected NO_x declines within these models trigger a more complete relaxation of titration, resulting in a larger O_3 increase.

The consistently strong positive O_3-TAS correlation, with all models exceeding 0.85 over China as a whole and in most sub-regions, underscores climate warming as a major and predominant driver of future O_3 trends. A clear exception is the Southwest China, where most models simulate lower correlations, reinforcing the interpretation that non-thermal controls (e.g., the complex terrain of the Tibetan Plateau and international transport) play a unique role there. Against this backdrop of high consensus, models like UKESM1-0-LL, which sustain correlations >0.9 across nearly all regions, likely simulate particularly efficient temperature-dependent processes (e.g., in BVOCs emissions or peroxy radical chemistry), thereby projecting an amplified warming effect on O_3.

Although a positive O_3-CH_4 correlation is a common feature across models, its highly variable strength highlights key uncertainties in how they represent methane's role. The MIROC-ES2H is a prominent outlier, simulating consistently weak and, in some regions even negative correlations ($r = -0.14$ in South China). This, coupled with its atypical weak (or even positive) O_3-NO_x correlation in the same region, strongly suggests that O_3 production in MIROC-ES2H is predominantly governed by the local interplay of NO_x and VOCs, whose strong control effectively masks the influence of the global CH_4 background. In contrast, models like IPSL-CM5A2 ($r = 0.51$ in China) simulate a clear positive correlation, indicating that their frameworks allow methane oxidation chemistry to play a substantive and detectable role in future O_3 production.

The strong but model-dependent O_3-VOCs positive correlations provide further diagnosis of chemical regime representations. Models such as EC-Earth3-AerChem and UKESM1-0-LL show high sensitivity to both NMVOCs and BVOCs emissions (with correlations often >0.8), which is characteristic of a VOC-sensitive chemical environment. A particularly telling diagnostic is that in these models, O_3 correlates more strongly with VOC emissions than with near-surface NO_x concentrations. This pattern potentially reflects differences in the represented chemical nonlinearity of the VOCs-NO_x-O_3 system or the relative atmospheric lifetimes and transport of the precursor species.

Q4. Conclusions - A similar issue to the abstract, which is reads like a list of results. It could be made more concise provide a more integrated discussion of the results and what this means in terms of performance and projection. Can you provide more of a discussion of how to take Biases learned in present day evaluation to historical and future projections?

Response: Thank you for your valuable comments. We fully agree that the original Conclusions section was overly focused on listing results. In response to your suggestions, we have thoroughly revised the section to significantly enhance its synthesis and discussion depth.

We have clarified this point in the revised manuscript Lines 920–968 as follows:

"This study provides a comprehensive evaluation of surface O_3 over China using the CMIP6 multi-model ensemble, spanning present-day performance, historical evolution (1850–2014), and future projections under diverse scenarios. By synthesizing findings across these temporal scales, we offer an integrated perspective on model capabilities, key sources of uncertainty, and the implications for understanding past changes and future air quality. The principal insights are discussed below.

(1) The CMIP6 MME reproduces the observed spatial and seasonal pattern of surface O_3, with maxima in summer and over central China. However, a systematic underestimation prevails across most models and regions, particularly in eastern China. This bias is not random; rather, it tends to be modulated by environmental conditions: model error generally increases over urban surfaces and under high $PM_{2.5}$ concentrations, revealing a complex, non-linear relationship especially in winter. Critically, these identified present-day biases are instructive for interpreting simulations of other periods. They point to persistent structural model limitations in key processes, particularly photochemical mechanisms and aerosol-chemistry interactions. Consequently, historical or future periods characterized by environmental conditions analogous to those linked with high bias today are likely subject to similar directional biases in simulated O_3. Acknowledging and accounting for this transferable understanding of model bias can therefore help advance a more robust interpretation of both past changes and future projections.

(2) The MME indicates a substantial increase of 39.3 ± 14.4 $\mu g \cdot m^{-3}$ in China's annual mean surface O_3 since 1850, with a marked acceleration post 1950. The large inter-model spread in this trend originates primarily from two factors highlighted in the literature: (i) differences in the simulated pre-industrial (1850) baseline concentrations, and (ii) divergent chemical sensitivities of O_3 production to emission changes. This spectrum of sensitivity is clearly evident in our results: MIROC-ES2H exhibits high sensitivity and simulates the largest historical trend, whereas UKESM1-0-LL shows lower sensitivity, producing a trend most consistent with a direct, emission-driven response. The fact that the greatest model disparity occurs in Central China, a region also undergoing significant aerosol changes, further highlights the contribution of aerosol-chemistry interactions to the overall projection uncertainty.

(3) Future surface O_3 levels over China are governed by the complex interplay between air pollution control and climate mitigation. A pronounced decrease occurs under the strong co-mitigation scenario SSP1-2.6. In stark contrast, the high-forcing SSP3-7.0 pathway projects continuous O_3 increases in many regions. Our analysis reveals that the common projection of a mid-century O_3 peak followed by a decline can arise from two fundamentally different policy dominances: pathways where climate mitigation objectives are central (SSP2-4.5), and pathways where improvements are primarily driven by stringent air quality controls even under high climate forcing (SSP5-8.5). This underscores a critical insight: substantial O_3 reductions are achievable under divergent socioeconomic futures, but they demand either deep coupled climate-air quality governance or, alternatively, exceptionally strong dedicated air pollution policies in the absence of climate action.

(4) This uncertainty arises from fundamental differences in how models represent key chemical and climatic processes. Under SSP3-7.0, while models generally agree on the positive roles of climate warming (TAS), global background chemistry (CH_4), and precursors (NMVOCs), they diverge sharply in simulating the local efficacy of NO_x control, as indicated by inconsistent O_3-NO_x correlations. Such divergence underscores that the simulated O_3 response is not merely a function of emission inputs but is critically dependent on each model's representation of complex, non-linear transition between chemical regimes (e.g., VOC-limited vs. NO_x-saturated), which can be further modulated by model-specific climate sensitivities and background emission levels. Consequently, the optimal strategy for future O_3 abatement, specifically, the relative priority and expected benefit of controlling NO_x versus managing other drivers like CH_4 or NMVOCs under a warming climate, remains highly model-dependent. This poses a fundamental challenge for deriving robust, model-consistent policy insights, highlighting the need to interpret projections within the context of these underlying chemical mechanisms.

Q5. Can the authors comments on the use of a multi-model mean and if any weight should be given to contributions from individual models? For example, UKESM1-0-LL and UKESM1-1-LL basically use the same the chemistry scheme so perhaps their results are giving too much weight to the MME. Also is there an issue in comparing results from experiments with a different number and type of model contributing to the MME (e.g. 9 models for SSP370 and 5 models for SSP126 or see line 535)?

Response: We thank the reviewer for the comment. In the revised manuscript, we have added a discussion at the end of the main text clarifying the use of the MME, the equal-weighting ("one-model-one-vote") approach, the potential dependence between models (e.g., UKESM1-0-LL and UKESM1-1-LL), and the differing number of models across scenarios. We also note the limitations of

the MME and suggest possible approaches to better quantify inter-model uncertainties in future work. We have clarified this point in the revised manuscript Lines 998–1008 as follows:

"Second, the MME was calculated using an equal-weighted scheme, following the common "one-model-one-vote" approach in model intercomparison studies. This method inherently assumes model independence. It should be noted that some models (e.g., UKESM1-0-LL and UKESM1-1-LL) share similar atmospheric chemistry schemes, which may bias the ensemble toward specific model families. Third, due to the varying data availability for different CMIP6 scenarios, the number of contributing models differs across scenarios. Consequently, the uncertainty ranges (e.g., SDs) of the MME are not strictly comparable between scenarios. Despite these limitations, the main qualitative conclusions of this study are robust, and the primary trends are not dominated by any single model. Future work could consider employing performance-based or independence-weighted ensemble methods, or utilizing more comprehensive and balanced model ensembles, to better quantify uncertainties arising from inter-model structural differences."

2. Minor Comments on Structure, Figures and Tables

(1) Suggestion of slightly re-wording title to: "Historical and Future changes of surface ozone over China from CMIP6 models, including an assessment of present-day uncertainties in model prediction."

Response: Thank you for the suggestion. The title has been revised accordingly.

(2) Should section 3.1 just be labelled as a present-day model evaluation section? I am not sure where temperature fits in (apart from the seasonal cycle)

Thank you very much for this helpful suggestion.

We agree that Section 3.1 mainly focuses on the evaluation of present-day conditions. Accordingly, we have revised the title of Section 3.1 to "Inter-model performance and spatiotemporal evaluation."

(3) Section 5 is very long, can this be split up in to different sub-sections to make the article read better?

Response: Thank you very much for this helpful suggestion. We agree that Section 5 was too long in the original manuscript. To improve readability, we have now divided Section 5 into several sub-sections:

5.1 Projected trends under different SSP scenarios

5.2 Model spread and its drivers under the SSP3-7.0 scenario

These structural adjustments make the section easier to follow.

(4) Table 1 – I think the UKESM1-1-LL reference should be https://gmd.copernicus.org/articles/16/1569/2023/ . Also could you include a few details on the models such as resolution here?

Response: Thank you for the comment. The reference for UKESM1-1-LL has been updated to: Mulcahy et al. (2023). In addition, details on the model resolutions have been added to Table 1 for clarity.

(5) Figure 1 – It is very hard to see the aerosol components on this chart so I would just not show them especially as this study is meant to focus on ozone. Also the number are also hard to read.

Response: We thank the reviewer for this comment. We agree that the original Figure 1 was visually cluttered, making the aerosol components and text hard to discern, and it did not directly support the ozone-focused narrative. To address this, we have replaced the original figure with a new, clear map illustrating the geographic locations of China and its seven sub-regions used throughout the study. This revision serves two purposes: (1) it eliminates the legibility issues by presenting essential spatial information in a clean, readable format; and (2) it provides a foundational reference for the regional analyses that follow, thereby strengthening the manuscript's focus.

The original analysis of O_3 and $PM_{2.5}$ based on TAP data has been removed, as a more detailed analysis using CMIP6 data is provided later in the manuscript.

Revised Figure 1 is shown below:

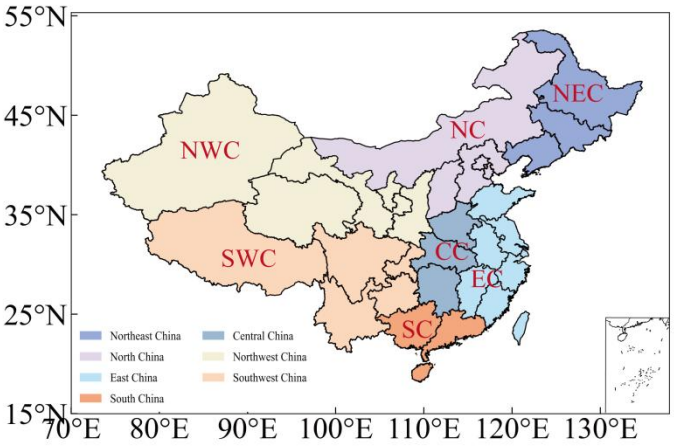

Figure 1. Geographic locations of China and its seven sub-regions used in this study (Northeast China (NEC), North China (NC), East China (EC), South China (SC), Central China (CC), Northwest China (NWC), and Southwest China (SWC)).

(6) Perhaps the statistics could be included in a table somewhere else? I wonder if it is worth including the CMIP6 multi-model mean for a direct comparison of temporal changes?

Response: We thank the reviewer for the suggestion. In response, we have added the relevant statistics in the Supplementary Materials: Table S2 summarizes the seasonal performance statistics of the multi-model ensemble (MME) O_3 relative to TAP, and Table S3 provides the annual correlation coefficients (r) between individual CMIP6 models and TAP. Additionally, in Figure 5 (Heatmap of O_3 concentration bias from nine CMIP6 models compared to TAP across different months, seasons, and regions in China and its sub-regions), we have included the MME values to facilitate a direct comparison of temporal changes among models.

Revised Figure and Tables are shown below:

Table S2. Summary of seasonal performance statistics of MME O_3 relative to TAP

Statistics ($\mu g \cdot m^{-3}$)	MAM	JJA	SON	DJF
Bias	-25.4	-12.8	-22.9	-25.3
SD (Inter-model)	16.8	18.8	15.2	19.0

Table S3. Annual correlation coefficients (r) between individual CMIP6 models and TAP.

Models	BCC-ESM1	CESM2-WACCM	EC-Earth3-AerChem	GFDL-ESM4	IPSL-CM5A2-INCA	MIROC-ES2H	MRI-ESM2-0	UKESM1-0-LL	UKESM1-1-LL	MME
r	0.78	0.91	0.77	0.80	0.80	0.88	0.81	0.73	0.74	0.80

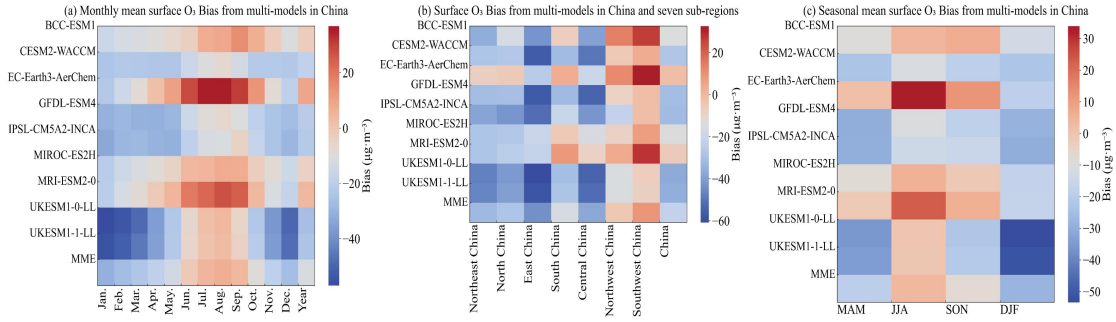

Figure 5. Heatmap of O_3 concentrations bias from nine CMIP6 models compared to TAP across different months, seasons and regions in China and sub-regions.

(7) Figure 2 – I might to suggest to use different colour bars for each column of plots that show different metrics as having the same colour scale for all is currently a bit confusing

Response: Thank you for this helpful suggestion. In the revised manuscript, we have updated Figure 2 by using separate colour bars for each column according to the specific metric being displayed. This modification improves visual clarity and avoids confusion caused by applying a single colour scale to different statistical metrics.

Revised Figure are shown below:

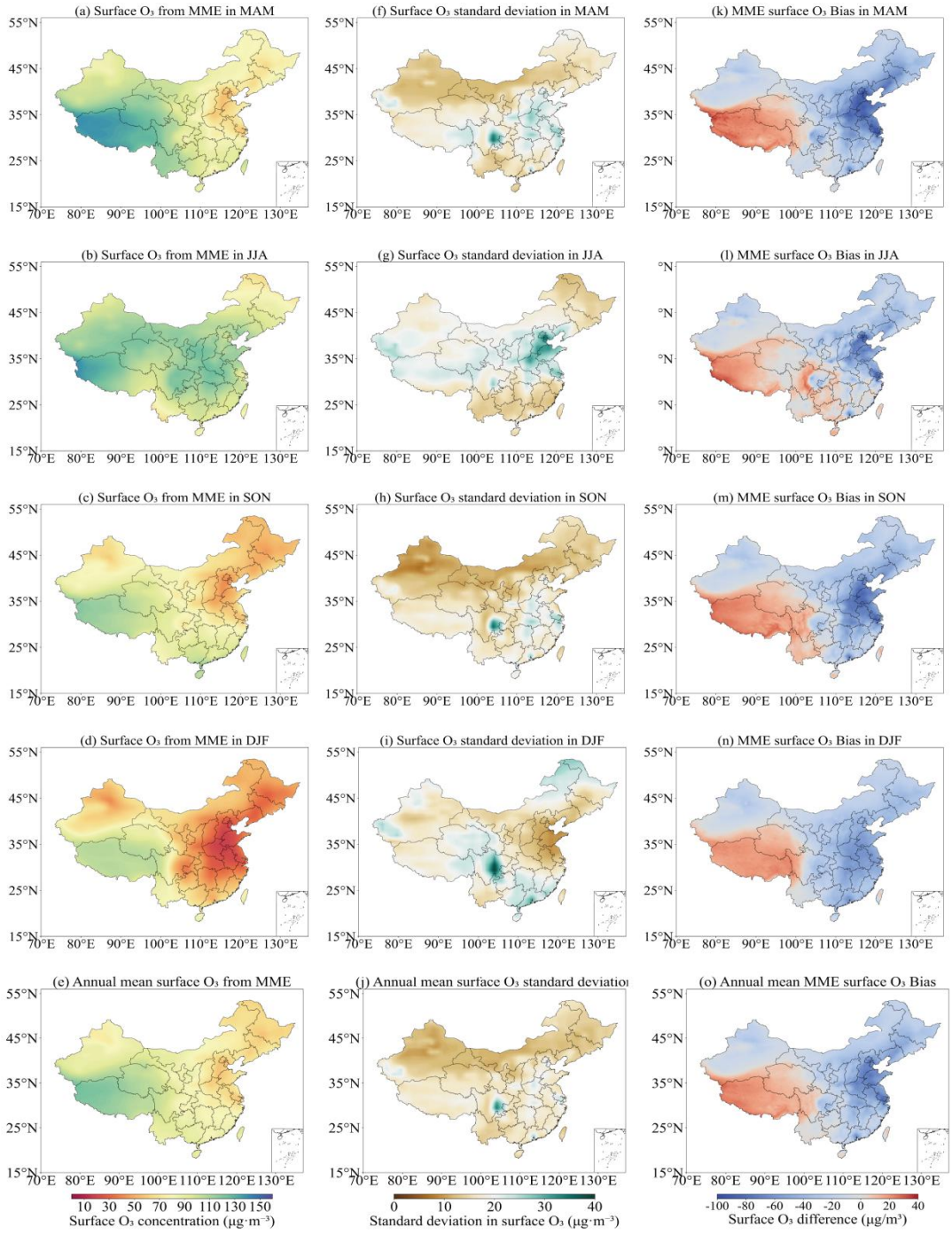

Figure 2. Multi-model (nine CMIP6 models) annual and seasonal mean surface O_3 concentrations over the 2015–2023 period in (a) MAM; (b) JJA; (c) SON; (d) DJF; and (e) annual mean. The SD of the MME in (f)MAM, (g) DJF, (h) JJA, (i)SON, and (j) annual mean. The difference between the MME and TAP observations in (k)MAM, (l) DJF, (m) JJA, (n)SON, and (o) annual mean.

(8) Figure 3 – perhaps the MME could be highlighted more for a direct comparison to TAP, with the individual models made to be more transparent or smaller.

Response: Thank you for the helpful suggestion. We agree that in the original Figure 3 the MME curve was not sufficiently distinguishable, which may have limited the clarity of the comparison with TAP. In the revised Figure 3, we have made the following improvements to enhance readability: (1) increased the line thickness and saturation of the MME curve, and (2) slightly reduced the opacity of the TAP shading. These adjustments improve the visual contrast and make the MME-TAP comparison clearer across all regions.

Revised Figure are shown below:

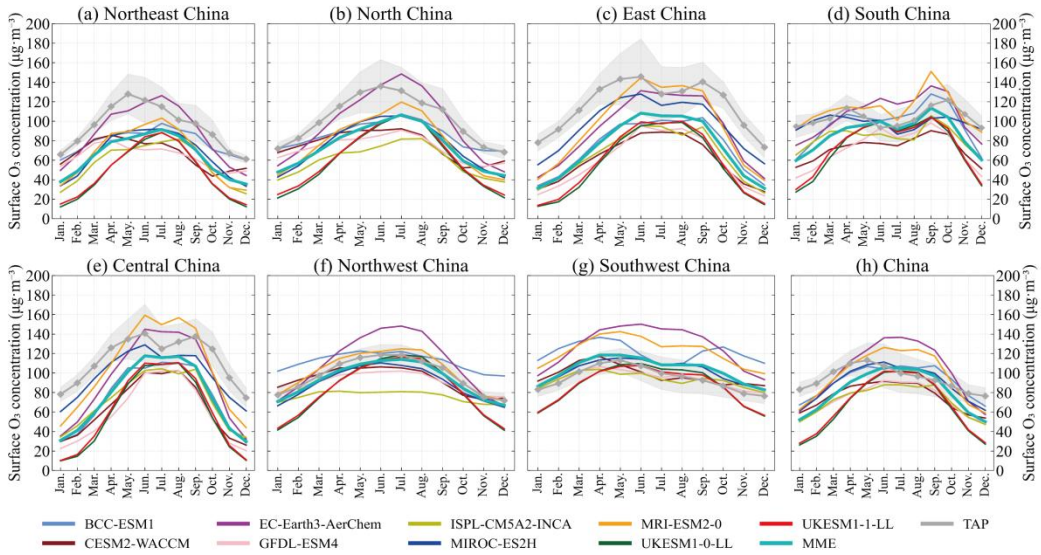

Figure 3. Comparison of the annual cycle of O_3 concentrations, between individual CMIP6 models, the MME and TAP in China and sub-regions for the period 2015–2023. The grey shading shows SD of TAP observations within the region.

(9) Figure 7 – More details in the caption on the Figure, what is (b) showing? and very similar colours to those on Figure(a) so confusing

Response: Thank you for the suggestion. We agree that the original caption of Figure 7 did not clearly describe the meaning of panel (b), and that the color scheme was too similar to panel (a), which may cause confusion.

In the revised manuscript, we have:

(1) Expanded the caption of Figure 7 to explicitly describe what panel (a) and (b) represents.

(2) Adjusted the color scheme of panel (b) to ensure clear visual distinction from panel (a).

The revised figure and caption have been updated accordingly.

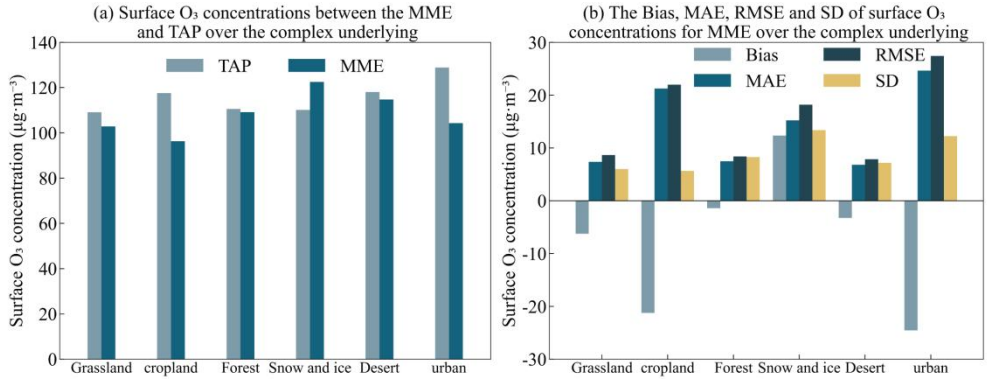

Figure 7. The bias, MAE, RMSE and SD of surface O_3 concentrations simulated by CMIP6 models relative to TAP from 2015 to 2023 over the complex underlying surface in China.

(10) Figure 8 – More details in the caption on the Figure

Response: Thank you for this comment. We agree that the caption of Figure 8 did not provide sufficient detail to clearly explain the content of each panel. In the revised manuscript, we have expanded the caption to describe the meaning of each sub-panel and the variables shown, so that the figure can be more easily understood without referring to the main text. The updated figure and caption have been included in the revised version.

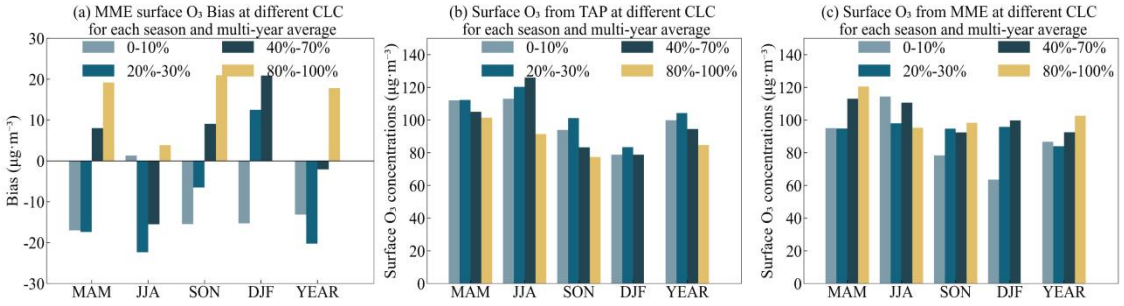

Figure 8. The bias in surface O_3 concentrations simulated by CMIP6 models relative to TAP from 2015 to 2023 under different total cloud cover levels in China.

(11) Figure10 – similar to another Figure, could the MME be made more prominent?

Response: Thank you for the helpful suggestion. This issue is consistent with the comment on Figure 3. In the revised Figure 10, we have similarly enhanced the visibility of the MME by increasing the line thickness and colour saturation, while reducing the opacity of the individual model curves. This improves the contrast and allows for a clearer comparison between the MME and TAP.

Revised Figure are shown below:

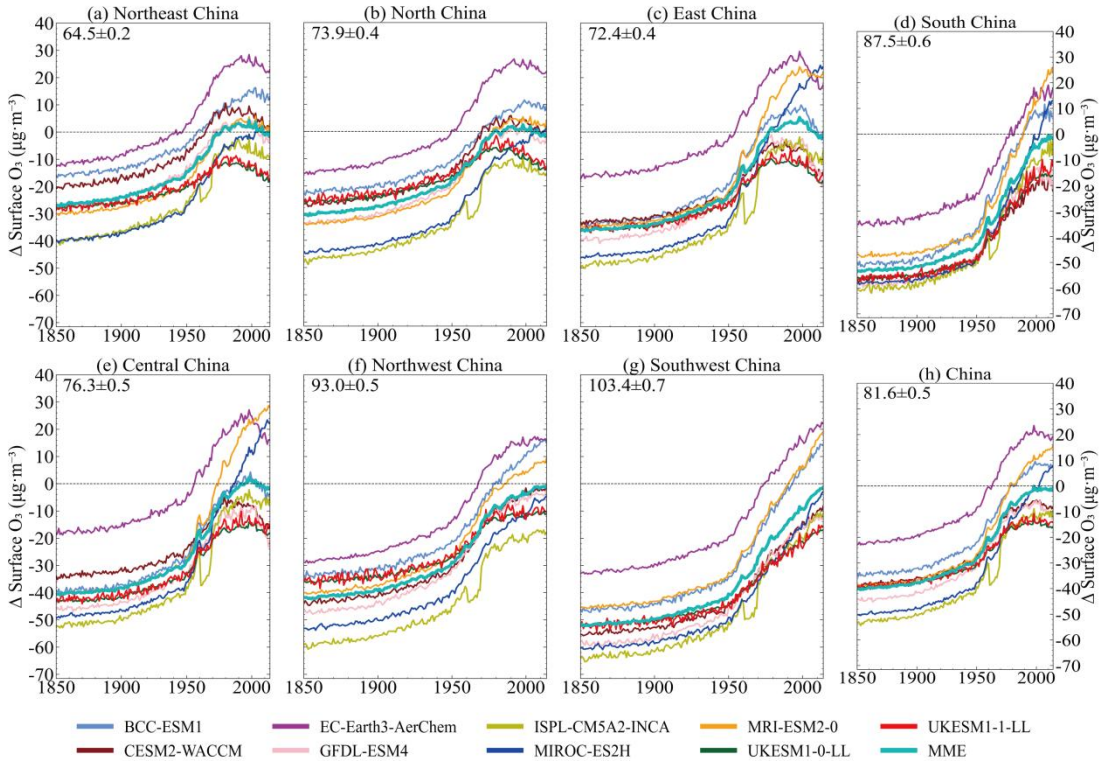

Figure 10. Changes in the China and sub-regions annual mean surface O_3 concentrations from the pre-industrial period to present day, relative to a 2015–2023 mean value, across nine CMIP6 models and MME. The multi-model annual mean 2015 – 2023 surface O_3 concentrations (±1 SD) are shown in the top left of each panel.

(12) Figure 11 – Why do some of the scenarios start about 30 μg m-3 above the zero line?

Response: Thank you for the comment. In Figure 11, the scenarios that start above +30 μg m^{-3} (SSP4-3.4, SSP4-6.0, and SSP5-3.4-OS) correspond to Tier 2 experiments in CMIP6, for which only one model (MRI-ESM2-0) provides ozone data. MRI-ESM2-0 simulates systematically higher present-day surface O_3 concentrations over China compared with most other CMIP6 models (second only to EC-Earth3-AerChem in our ensemble; see Figure 3 and Figure S2). Because Figure 11 presents deviations relative to the 2015–2023 multi-model mean, the higher absolute O_3 levels in MRI-ESM2-0 result in higher starting deviations for these Tier 2 scenarios. To avoid confusion, we have clarified this in the figure caption as follows (Lines 715–720) :

"Figure 11. Future China and sub-regions changes in annual mean surface O_3 under different CMIP6 SSP scenarios. The dashed black line represents the curve of the difference at zero. For the Tier 2 scenarios (SSP4-3.4, SSP4-6.0, and SSP5-3.4-OS), O_3 data are available only from MRI-ESM2-0. As MRI-ESM2-0 simulates higher O_3 over most regions of China compared to the MME, these scenarios

show larger initial deviations in the national panel and in some regions. The multi-model regional mean for 2015–2023 is shown in the top-left corner of each panel."

(13) Figure 15 – Could you make better use of this of this Figure to help explain where future changes sit in terms of NOx/VOC ratio?

Response: Thank you for the comment. The O_3 formation regime shown in Figure 15 is derived from present-day observations and should be regarded as a diagnostic reference rather than a fixed projection of future atmospheric chemistry, which may evolve with changes in precursor emissions. This is also indicated in the caption of Figure 15: "Distribution of O_3 concentration in the NO_2-NMVOCs coordinate over China, based on 2015–2023 surface observations of NO_2 and O_3 and MME NMVOCs under the SSP3-7.0 scenario." This point has been clarified in the revised manuscript as follows (Lines 831–844).

"It is important to note that the O_3 formation regime depicted in Figure 15 is derived from present-day observations and should be regarded as a diagnostic reference rather than a fixed projection of future atmospheric chemistry. CMIP6 multi-model simulations under SSP3-7.0 project that NO_x emissions increase from 2015 and peak around 2035 before gradually declining to near 2015 levels by 2065(Figure S3), while NMVOCs emissions continue a modest upward trend throughout the century (Figure S6). Based on these projections, our analysis suggests that many regions in China could remain in or near the high-NO_x suppression regime during the first half of the 21st century (around the NO_x emission peak), where O_3 formation could be inhibited by excessive NO_x. As NOx emissions decline thereafter, the chemical environment is projected to gradually shift towards a regime more sensitive to NMVOCs, potentially approaching VOC-limited conditions, where O_3 production becomes increasingly dependent on NMVOCs levels. Regional variations in the relative reduction rates of NO_x and NMVOCs are expected to further influence these transitions. These potential shifts highlight the need for adaptive, region-specific emission-control strategies that account for evolving chemical environments."

3. Minor Comments on Text

(1) Line 16 – replace "current" with "present-day"

Response: We appreciate the reviewer's careful reading. The term has been revised as suggested in Line 17.

(2) Line 35 – insert "gas" between "trace" and "components"

Response: We appreciate the reviewer's careful reading. The term has been revised as suggested in Line 40.

(3) Line 35-39 – Sentence is quite long. Could be broken up into two parts to read better

Response: Thank you very much for the helpful suggestion. We have rewritten the long sentence and split it into shorter ones to improve readability. Please see Lines 40–44 in the revised manuscript.

The specific content is " Ozone (O_3) is one of the most important trace gases components in the Earth's atmosphere, serving as a protective barrier for the global ecosystem and a crucial heat source in the stratosphere (Zhang et al., 2017). Variations in ozone concentrations strongly influence the climate of both the stratosphere and the troposphere (Xie et al., 2017; Haase and Matthes, 2019; Lin and Ming, 2021), and play a critical role in controlling the atmospheric temperature structure."

(4) Line 50 – I am not sure "organisms" is the right word here. Maybe replace with "ecosystems"?

Response: Thank you for your suggestion. We have replaced "organisms" with "ecosystems" to improve accuracy. Please see Line 55 in the revised manuscript.

(5) Line 51 – "past few years" – can you be more specific about the time period?

Response: Thank you for the suggestion. We have revised the wording accordingly. In Line 56 of the article, the phrase "Over the past few years" has been replaced with "Over the past decade" to more clearly indicate the referenced time period.

(6) Line 76 -77 – Are there are more recent observations to show the continual decline in Air Quality due to increasing ozone? What time period are you referring to in this sentence?

Response: Thank you for raising this important point. We have revised the statement to clearly define the time period and provide up-to-date observational support.

The revised sentence in Lines 80–81 of the manuscript now reads:

"Against this backdrop, surface O_3 pollution in China continues to worsen and expand from 2015 to 2024 (Wang et al., 2020; Xiao et al., 2022; Figure S8)."

To support this trend, we have added recent observational studies (Wang et al., 2020; Xiao et al., 2022) and clarified the reference years in the revised sentence (Lines 80–81). In addition, Supplementary

Figure S8 now presents nationwide changes in pollutant concentrations and the exceedance-day ratio based on 1181 monitoring stations, further confirming the continuous rise in O_3 pollution."

(7) Line 80 – What is the "Dual-Carbon" strategy and how does this relate to air pollutants since it seems to be referring to carbon?

Response: Thank you for raising this important point. The "Dual-Carbon" strategy refers to China's national goals of achieving a carbon emissions peak by 2030 and carbon neutrality by 2060. The relevance to air pollutants lies in the synergistic control of greenhouse gases and air pollutants. The primary sources of CO_2 (e.g., power, industry, transport) are also major emitters of ozone precursors like NOx and VOCs. Therefore, measures to reduce fossil fuel combustion under this strategy will have co-benefits for mitigating both climate change and ozone pollution.

We have revised the introduction (Lines 84–92) to explicitly include this explanation, the revised sentence now reads:

"Under the guidance of China's national "Dual-Carbon" strategy, which targets carbon peaking by 2030 and carbon neutrality by 2060, synergistic control of greenhouse gas emissions and air pollutants has become essential for achieving sustained improvements in air quality. This is because the dominant CO_2 emission sectors (e.g., power generation, industrial production, and transportation) are also major sources of ozone precursors such as NO_x and anthropogenic VOCs. Therefore, reducing fossil fuel combustion can bring co-benefits by mitigating both climate warming and O_3 pollution. Given this context, accurately understanding the spatiotemporal evolution of surface O_3 concentrations in China is of great importance, as it provides a scientific basis for evaluating the effectiveness of emission control measures and for guiding future air quality management strategies."

(8) Line 92 – What about other models such as Chemical Transport Models (CTMs), are these not also used as well?

Response: We thank the reviewer for this comment. We have added a clarifying sentence in the manuscript (Lines 101–106) to indicate that Chemical Transport Models (CTMs) can also be used for O_3 research, and to explain their distinction from Chemical-Climate Models (CCMs):

"For decades, both chemical-climate models (CCMs) and chemical transport models (CTMs) have been indispensable tools for studying global surface O_3. CTMs are widely used for high-resolution regional simulations, while CCMs are particularly suited for investigating long-term global trends, reproducing historical O_3 distributions, predicting future changes across various spatial and temporal scales, and exploring interactions with precursors and atmospheric physical-dynamic processes."

(9) Line 101 – Linked to the above, perhaps at this stage it might be worth mentioning other multimodel initiatives that have attempted to understand surface ozone such as HTAP (Hemispheric Transport of Air Pollutants) and its regional counter-part MICS-Asia (https://acp.copernicus.org/articles/special_issue390.html). Additionally some comments on the TOAR (Tropospheric Ozone Assessment Report) multi-model comparison would be useful.

Response: Thank you for this valuable suggestion. We have added a paragraph in the revised manuscript (Lines 113–118) introducing key multimodel initiatives such as HTAP, MICS-Asia, and TOAR, and briefly describing their contributions to understanding surface O_3 variability:

"Beyond individual model studies, international multimodel initiatives have advanced understanding of surface O_3 pollution. The Hemispheric Transport of Air Pollution (HTAP) project (Fiore et al., 2009) and its regional counterpart MICS-Asia underscore the role of hemispheric and transboundary transport in shaping regional O_3 patterns (Li et al., 2019). The Tropospheric O_3 Assessment Report (TOAR) provides a global benchmark for evaluating surface O_3 models and assessing historical trends using harmonized observations and multimodel comparisons (Young et al., 2018)."

(10) Line 113 – Include "in the present-day." After "current uncertainties"

Response: The expression "in the present-day" has now been added in Line 130 of the revised manuscript.

(11) Line 171 – Are these average values for the full 24 years of the TAP data? If so is this useful given the large changes in both O_3 and $PM_{2.5}$ that have occurred over this period?

Response: We thank the reviewer for this helpful comment. Yes, the values shown were averages over the full 24-year TAP record. We agree that such long-term averaging may obscure substantial interannual changes in both surface O_3 and $PM_{2.5}$ over this period, particularly given the pronounced emission trends. Given this limitation, and considering that the TAP-based analysis was intended only as contextual background to serve as an observational reference for comparison with CMIP6 simulations and to provide a qualitative assessment of model uncertainty, we have removed the TAP-based averaged O_3 and $PM_{2.5}$ results and their associated discussion from the revised manuscript, we have removed the TAP-based averaged O_3 and $PM_{2.5}$ results and their associated discussion from the revised manuscript. The figure is now retained only to show the geographic locations of China and its seven sub-regions used in this study, which provides essential spatial context for the subsequent analyses. The main quantitative assessment of surface O_3 is based on CMIP6 multi-model simulations and is presented in later sections.

(12) Line 198 – also the largest standard deviation as well

Response: We thank the reviewer for this comment. We have revised the text to explicitly clarify the relationship between seasonal mean O_3 concentrations and inter-model variability, and added the corresponding seasonal standard deviation values in Lines 292–295 and Table S2.

The specific content is:

"The inter-model SD (Table S2) is largest in DJF (19.0 $\mu g \cdot m^{-3}$), but this peak is mainly confined to the Sichuan Basin in Southwest China and parts of South China. In contrast, although the SD in JJA is slightly lower (18.8 $\mu g \cdot m^{-3}$), elevated values extend across a much broader area, particularly in North China and East China. The SD decreases in MAM (16.8 $\mu g \cdot m^{-3}$) and is smallest in SON (15.2 $\mu g \cdot m^{-3}$)."

(13) Line 213 – So does this imply that the TAP data has been made with a stronger southeast Monsoon? Where has the meteorological data come from?

Response: We thank the reviewer for the question. The TAP dataset does not include meteorological fields that allow us to diagnose monsoon strength. Our explanation relies on previous studies of southeast monsoon behavior (e.g., Yin et al., 2019), rather than on meteorological information within the TAP dataset. We have clarified this point in the revised manuscript.

Revised text in Lines 250–253:

"This seasonal pattern may be related to the southeast monsoon, which has been shown in previous studies to strengthen during JJA and enhance northward transport while increasing precipitation and humidity (Yin et al., 2019), thereby reducing O_3 concentrations."

(14) Line 217 – What about source of ozone from the free troposphere and stratosphere that could influence this region?

Response: Thank you for the valuable comment. Previous studies have shown that, in addition to the transboundary transport from South Asia, O_3 in Southwest China — especially over the Tibetan Plateau — is also influenced by large-scale background contributions from the free troposphere and the lower stratosphere. We have added the following text in the revised manuscript (Lines 257–264):

"In addition, recent modeling work has shown that the Tibetan Plateau receives substantial background O_3 from the free troposphere and long-range transport. Li et al. (2014) demonstrated that trans-Eurasian transport contributes 10–15 ppbv of surface O_3 to western China, superimposed on a natural background of 35–40 ppbv, with particularly strong influence from Indian emissions during the summer monsoon. Furthermore, the Tibetan Plateau is a known hotspot of stratosphere–troposphere

exchange (STE), where stratospheric ozone can be transported downward into the troposphere through tropopause folding and subsidence, further contributing to elevated background O_3 levels in this region (Yin et al., 2023)."

(15) Line 221 - Why is ozone always so high is Northwest China along with no seasonal cycle? Is this because of lower local emissions and higher import from extra-regional sources?

Response: Thank you for the valuable comment. In Northwest China, surface O_3 concentrations are high throughout the year, with a seasonal cycle that peaks in JJA. The persistently elevated O_3 levels result from multiple factors, We have added an explanation in the revised manuscript (Lines 264–278):

"In Northwest China, O_3 concentrations remain high across all four seasons, although a seasonal cycle is present with the highest values in JJA (113 $\mu g \cdot m^{-3}$).The high O_3 levels can be attributed to multiple factors. First, the region is arid and receives strong solar radiation, which promotes efficient photochemical O_3 production. Second, although overall anthropogenic emissions are relatively low compared with eastern and southern China, some cities such as Urumqi and Yinchuan still exhibit relatively high O_3 concentrations during summer, indicating that local emissions from industrial, energy, and traffic activities cannot be neglected (Zhu et al., 2023). Third, Northwest China can be affected by long-range transport of O_3 and its precursors from surrounding and remote source regions. Modeling studies have shown that European anthropogenic emissions can enhance surface O_3 over northwestern China by approximately 2–6 ppbv during spring and summer, while contributions from the Middle East may occasionally add another 1–4 ppbv (Li et al., 2014). Finally, high-altitude areas are more susceptible to STE, allowing O_3 from the upper troposphere and lower stratosphere to be transported downward, further elevating surface O_3 background levels (Yin et al., 2023)."

(16) Line 223 – Similarly why is ozone much lower in north east China?

Response: Thank you for the insightful comment. We have added an explanation for the relatively low O_3 concentrations in northeast China in the revised manuscript (Lines 278–286):

"Northeast China, on the other hand, has the lowest O_3 concentrations among all sub-regions, particularly during MAM, JJA, and SON, with concentrations of only 76 $\mu g \cdot m^{-3}$, 88 $\mu g \cdot m^{-3}$, and 54 $\mu g \cdot m^{-3}$, respectively. This can be attributed to several factors: first, the cooler climate and shorter period of strong solar radiation in northeast China slow down the rate of photochemical ozone production. Second, the region has relatively lower emissions of reactive VOCs compared to the more industrialized and densely populated regions of eastern and southern China, which constrains the overall ozone production potential. Third, northeast China is frequently influenced by clean continental

air masses from Siberia, which effectively dilute local and regional pollutant concentrations."

(17) Line 229-230 – I would say that the SD is only largest in DJF in in Sichuan basin. I would say there is larger SD across more regions in JJA.

Response: Thank you very much for the insightful comment. We agree with your observation and have revised the corresponding sentence to clarify this spatial difference in Lines 293–296 as follows:

"The inter-model SD (Table S2) is largest in DJF (19.0 $\mu g \cdot m^{-3}$), but this peak is mainly confined to the Sichuan Basin in Southwest China and parts of South China. In contrast, although the SD in JJA is slightly lower (18.8 $\mu g \cdot m^{-3}$), elevated values extend across a much broader area, particularly in North China and East China. The SD decreases in MAM (16.8 $\mu g \cdot m^{-3}$) and is smallest in SON (15.2 $\mu g \cdot m^{-3}$)."

(18) Line 235 – Potentially larger biases in MAM and SON?

Response: Thank you for the insightful comment. We have addressed this by providing additional details in Lines 301–308 as follows:

"The spatial distribution of the MME bias reveals distinct seasonal and regional patterns. On a national average, MME underestimate surface O_3 concentrations, with the largest negative biases occurring in MAM (-25.4 $\mu g \cdot m^{-3}$) and DJF(-25.3 $\mu g \cdot m^{-3}$), and smaller ones in SON(-22.9 $\mu g \cdot m^{-3}$) and JJA (-12.8 $\mu g \cdot m^{-3}$). Spatially, the most pronounced underestimations are found over eastern and northern China, particularly during MAM and SON. In contrast, positive biases (overestimations) are evident in Southwest China and in parts of Northwest China (primarily Qinghai Province), especially during SON. Notably, Northwest China (excluding the Qinghai overestimation) exhibits the smallest absolute biases among all sub-regions, with simulated O_3 concentrations aligning most closely with the TAP data across all seasons."

(19) Line 237 – It maybe worth reporting all the statistics (not just the bias ones here) in a table for easy reference to the reading. Same with the correlations on Line 246

Response: Thank you for the constructive suggestion. We have added a table summarizing all relevant statistics, including bias, standard deviation (inter-model), and correlation coefficients (Table S2-S3), to facilitate easier reference for readers.

(20) Line 249 – I think delayed maximum ozone is true in some regions (North China) but perhaps more similar in others (Northwest China)

Response: Thank you for the valuable comment. We have revised the manuscript to more clearly

describe the regional differences in the timing of the O_3 peak, including the specific peak months for each region. The revised text has been added in Lines 319–322 as follows:

"(1) The timing of O_3 peak concentrations in the MME (mostly in July–August, except for East and Central China in June, Southwest China in April) is overall slightly delayed compared to TAP (mostly in May–June), which is consistent with the results of the ACCMIP models (Young et al., 2018)."

(21) Line 250-252 – It is interesting that an underestimation of ozone concentrations is reported here whereas, in Figure 2 it seems like outside of Eastern China there is a lot of overestimation of ozone. Perhaps the MME could be made clearly on Figure 3 so that the under/over estimation across regions is clearer.

Response: We thank the reviewer for this insightful comment. The apparent inconsistency indeed stems mainly from the color scale used in Figure 2. In the bias plots, red colors do not necessarily represent positive biases. Because the color bar ranges from -100 to +20 μg m^{-3}, light red also corresponds to small negative biases, and only values above 0 indicate overestimation. To avoid this misunderstanding, we have revised the caption of Figure 2 to explicitly clarify the meaning of the color scale and have enhanced the color-bar labeling to highlight the zero-bias level.

We also agree that the MME curve in Figure 3 was not sufficiently distinguishable, which may have made the regional underestimation/overestimation patterns less clear. In the revised Figure 3, we have (1) increased the line thickness and saturation of the MME curve, and (2) slightly reduced the opacity of the TAP shading. These adjustments improve the visual contrast and make the MME-TAP comparison clearer across all regions.

Overall, these revisions help demonstrate more clearly that the MME exhibits underestimation across most regions, consistent with the main text. The related descriptions in the manuscript have been slightly revised accordingly (see Lines 322–328) as follows:

"The nine CMIP6 models evaluated in this study generally underestimate O_3 concentrations across most sub-regions of China. The most pronounced underestimations occur in UKESM1-0-LL and UKESM1-1-LL. Although these two models show slight overestimations in Southwest China from June to September, as well as in South China in June and in Northwest China in August, underestimation remains the dominant pattern, particularly in East and Central China during MAM and DJF, where simulated O_3 concentrations are over 60 μg·m^{-3} lower than the TAP values."

(22) Line 252 – the biases in UKESM1 are explored further in this paper which might help explain things https://acp.copernicus.org/articles/22/12543/2022/

Response: We thank the reviewer for the suggestion. We have added Liu et al. (2022) to further support the discussion of UKESM1 biases and updated the manuscript accordingly (Lines 328–334) as follows:

"This is consistent with the findings of Turnock et al. (2020), who suggested that these biases may be related to excessive NO_x titration in the UKESM1-0-LL model, leading to an underestimation of surface O_3 concentrations over much of the Northern Hemisphere continental regions during DJF. Liu et al. (2022) further conducted a global evaluation of surface O_3 biases in UKESM1, identifying temperature, photolysis rates, and regional chemical composition as key predictors, which may help explain the underestimations identified in this study."

(23) Line 289 – Are the biases in the MME for different seasons shown? If not could they be included on Figure 5?

Response: We thank the reviewer for the suggestion. To clearly show the seasonal performance of the MME, we have added the seasonal mean bias (DJF, MAM, JJA, SON) to Figure 5(c).

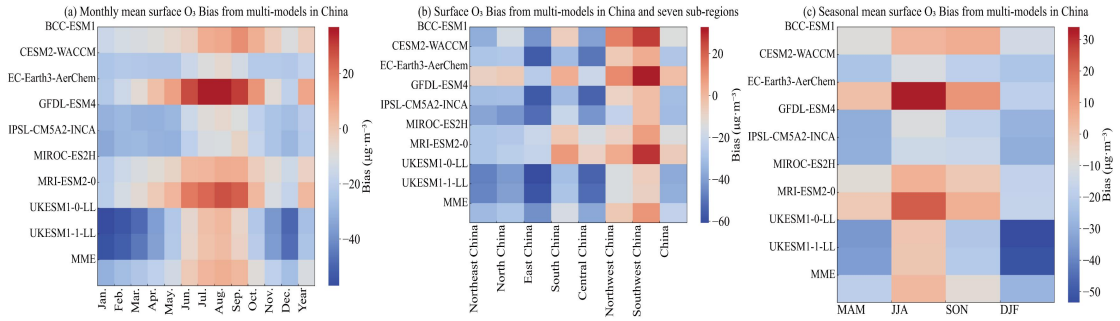

Figure 5. Heatmap of O_3 concentrations bias from nine CMIP6 models compared to TAP across different months, seasons and regions in China and sub-regions

(24) Line 325 – Do the natural land surface have consistent physical and chemical properties with less human influence? What about deforestation on forests and BVOCs/Fires which can mean quite complicated processes.

Response: We thank the reviewer for this insightful comment. We agree that natural land surfaces can involve complex physical and chemical processes, especially in regions influenced by forest loss, BVOC-rich vegetation, or fire activity. In our study, however, the analysis focuses on Northwest China during JJA, where the dominant natural land types (grassland, desert, forest, and perennial snow/ice)

generally experience less intensive land-use change and weaker direct human disturbance compared with cropland and urban surfaces. Under these conditions, biogenic emission patterns tend to vary less, which likely contributes to the smaller bias, MAE, and RMSE observed for natural land surfaces in the MME simulations. To clarify this point, we have revised the text in the manuscript (Lines 400–418) as follows:

"Natural land surfaces tend to provide more consistent physical and chemical conditions, with limited direct human interference, leading to relatively more accurate simulations of surface O_3 concentrations. Although natural ecosystems can involve complex processes, such as BVOCs emissions from vegetation, deforestation, or wildfires, our analysis based on long-term (JJA) MME effectively smooths out the random uncertainties associated with such transient events. Consequently, the underlying stability of natural surfaces in our study region becomes the dominant feature. In Northwest China during JJA the dominant natural land types (grassland, desert, forest, and snow and ice) generally undergo limited land-use change and exhibit weaker anthropogenic disturbances compared with cropland and urban surfaces. Under these relatively stable conditions, biogenic emission patterns are more predictable, as demonstrated by global modeling studies using MEGAN2.1 (Guenther et al., 2012). In addition, dry deposition, a major sink of tropospheric O_3, is more reliably captured over natural surfaces, where vegetation cover and leaf area index are relatively stable. Accurate parameterization of both stomatal and non-stomatal deposition pathways ensures reliable O_3 removal, reducing bias, MAE, and RMSE in multi-model ensemble simulations (Val Martin et al., 2014). In contrast, urban surfaces, characterized by intensive human activities and diverse anthropogenic emission sources (e.g., transportation and industry), present a more complex environment for O_3 formation and destruction. The heterogeneous land cover complicates dry deposition parameterization, often leading to underestimation of O_3 removal and larger discrepancies in simulated concentrations."

(25) Line 329 – What about the impact of deposition on land surfaces here?

Response: We thank the reviewer for this insightful comment. The influence of dry deposition on land surfaces has now been clarified in the revised manuscript. This clarification has been incorporated into the same updated section described in our response to Comment (24).

(26) Line 335-339 – If the models do not account for accurate deposition on snow/ice is it worth this much detail and is this a big issue over China?

Response: We appreciate the reviewer's insightful comment. We retain a brief discussion of snow/ice deposition because northern China experiences extensive and persistent snow cover during winter and spring, during which uncertainties in deposition parameterizations can influence cold season O_3

simulations and may have broader climatological implications. We have shortened and streamlined this section. The revised text has been added in Lines 419–431 as follows:

"Snow and ice surfaces can influence O_3 dry deposition during the cold season, particularly in northern China where snow cover often persists for several months. Their high albedo reduces the absorption of solar radiation at the surface and weakens photochemical activity, including the photolysis of NO_2 that contributes to O_3 formation. In addition, snow and ice generally exhibit lower O_3 uptake capacity, with deposition velocities (0.03 $cm \cdot s^{-1}$) typically smaller than those over vegetated or soil surfaces (Wesely et al., 1981). The magnitude and direction of O_3 flux over snow also depend on factors such as deposited trace gases, solar irradiance, snow temperature, and the underlying substrate (Helmig et al., 2007), but current atmospheric chemistry models often lack detailed parameterizations for these processes. As a result, O_3 deposition over snow and ice remains a potential source of uncertainty for cold-season simulations. It should also be noted that, in the context of O_3 removal, wet deposition contributes minimally to O_3 removal because of its low solubility. Therefore, while dry deposition over snow and ice may be locally important in winter or high-altitude regions, wet deposition has little impact on surface O_3 concentrations."

(27) Line 364-365 – Is wet deposition a large removal term for ozone?

Response: We appreciate the reviewer's comment. Wet deposition contributes minimally to O_3 removal due to its low solubility in water. We have added this clarification in the revised manuscript within the section updated in response to Comment (26).

(28) Line 366 – Which precursors?

Response: Thank you for the comment. We agree that wet deposition does not efficiently remove the primary O_3 precursors (NOx and VOCs). To avoid confusion, we have removed the previous wording referring to the removal of precursors from the revised manuscript. The modifications can be found in Lines 468–468 as follows:

"This pattern is primarily attributed to the attenuation of solar radiation by clouds , which suppresses photochemical O_3 production. In addition, accompanying precipitation processes can influence atmospheric stability and boundary layer development, which may indirectly affect surface O_3 concentrations."

(29) Line 380-384 – Isn't this what the models try to do? Can you be a bit more specific as this statement is very general.

Response: Thank you for the comment. We have revised the text to make it more specific regarding the mechanisms that need attention in CMIP6 model simulations. The modifications can be found in Lines 480–487 as follows:

"Therefore, when using CMIP6 models for O_3-related assessment and projection, it is important to consider how cloud cover modulates shortwave radiation and photochemical reaction rates, how precipitation affects atmospheric stability and boundary layer development, and how other meteorological variables such as temperature, humidity, and wind influence both chemical reactions and pollutant transport. Proper attention to these interactions is particularly crucial under polluted and complex meteorological conditions, where the parameterization of these coupled processes in current models remains challenging. Improving their representation is therefore key to reducing systematic uncertainties and enhancing projection accuracy (Jacob and Winner, 2009)."

(30) Line 390-393 – Can you be more specific about how aerosols alter the chemical composition?

Response: Thank you for the comment. We have revised the text to specify how aerosols affect O_3 formation. The changes can be found in Lines 493–499 as follows:

"Aerosols play a crucial role in the simulation of surface O_3 concentrations. Variations in $PM_{2.5}$ can influence O_3 formation through several pathways: (1) chemical effects: changes in aerosol composition (e.g., sulfate, nitrate, organic matter, BC) can perturb the concentrations of key radicals (e.g., OH, HO_2) and thus alter photochemical reaction rates (Wang et al., 2019); (2) radiative effects: aerosols modify optical properties and reduce actinic flux at the surface, thereby suppressing photolysis rates and slowing O_3 production; and (3) heterogeneous effects: reactions on aerosol surfaces can directly consume NO_x or other precursors, shifting the chemical environment for O_3 formation (Qu et al., 2021)."

(31) Line 401 – so this is suggesting that the aerosol impact on ozone formation is smaller in summer/autumn?

Response: Thank you for this insightful comment. Our results suggest that the net inhibitory effect of aerosols ($PM_{2.5}$) on surface O_3 formation is indeed generally smaller in summer (JJA) and autumn (SON) compared to winter (DJF), which aligns with established understanding in atmospheric chemistry. We have revised the text to clarify the underlying mechanisms. The modifications can be

found in Lines 502–523 as follows:

"The results show that during the JJA and SON, when $PM_{2.5}$ concentrations are relatively low, TAP data indicate an increase in O_3 concentrations with rising $PM_{2.5}$ levels. However, during the DJF, when $PM_{2.5}$ concentrations are higher, O_3 concentrations decrease as $PM_{2.5}$ levels increase. This is primarily due to the fact that, during JJA and SON, although the increases in $PM_{2.5}$ concentrations can attenuate solar radiation and participate in heterogeneous uptake of HO_2 and RO_2 radicals. However, their overall inhibitory effect on O_3 formation is relatively weaker during these seasons, because abundant sunlight and strong photochemical activity compensate for aerosol-induced reductions in photolysis rates. In addition, O_3 formation is typically in a VOC-limited regime, under which heterogeneous radical loss plays a smaller relative role (Liu et al., 2019). In contrast, during DJF, limited solar radiation, stronger atmospheric stability, and elevated NO_x concentrations amplify the aerosol impacts, reduced photolysis rates become more critical, heterogeneous radical uptake more effectively suppresses the HO_x cycle, and NO_x titration is more likely to occur. Meanwhile, the increase in BC concentration in $PM_{2.5}$ (Figure S1) enhances the light absorption of $PM_{2.5}$, further reducing UV radiation and inhibiting O_3 photochemical production. Additionally, other components of $PM_{2.5}$, such as NO_3^- and OM, may influence O_3 through specific pathways, particulate nitrate photolysis generates NO_2 and OH radicals, promoting O_3 formation, while OM and organic nitrates modulate radical (RO_x/NO_x) cycling and NO_x partitioning, thereby affecting O_3 photochemistry (Li et al., 2025). Observational evidence from Chen et al. (2021) in Delhi supports this seasonal behavior, reductions in $PM_{2.5}$ significantly enhance solar radiation, reduce AOD, and increase the photolysis rate of key O_3 precursors, thereby exacerbating O_3 accumulation under VOC-limited conditions, particularly in winter, when AOD is reduced by 50%, O_3 generation can increase by 25%."

(32) Line 406-407 – which pathways?

Response: We thank the reviewer for this comment. The pathways through which other $PM_{2.5}$ components, such as nitrate (NO_3^-) and organic matter (OM), may affect O_3 concentrations have been added to the revised manuscript (Lines 516–519).

"Additionally, other components of $PM_{2.5}$, such as NO3- and OM, may influence O3 through specific pathways, particulate nitrate photolysis generates NO2 and OH radicals, promoting O3 formation, while OM and organic nitrates modulate radical (ROx/NOx) cycling and NOx partitioning, thereby affecting O3 photochemistry (Li et al., 2025)."

(33) Line 464 – "driven solely by precursor emission variations" – I am not sure I understand this point

Response: We thank the reviewer for this helpful comment. To improve clarity, we have revised the sentence to better explain the intended meaning. The updated text can be found in Lines 578–583 as follows:

"Moreover, the simulated historical changes in surface O_3 from UKESM1-0-LL are comparable to those from the HTAP_param (an emission-only-driven model), indicating that the magnitude of change simulated by UKESM1-0-LL is similar to that solely from changes in precursor emissions (Turnock et al., 2020). This implies that UKESM1-0-LL may exhibit relatively weak contributions from non-emission-driven processes (e.g., chemistry-climate interactions) compared with other CMIP6 models."

(34) Line 471-473 – yes this might be an impact but can you say what time period this might be more important over?

Response: We thank the reviewer for this comment. To clarify the temporal context, we have revised the text in Lines 593–596 of the manuscript as follows:

"This aerosol modulation effect is likely most pronounced during seasons characterized by high $PM_{2.5}$ loadings and weak photochemical activity (e.g., winter), when reduced solar radiation and enhanced atmospheric stability may amplify the impact of aerosol-related processes on surface O_3."

(35) Line 482 – could also present percentage changes here too?

Response: We thank the reviewer for this comment. we have increased the present percentage changes in Lines 620–623 of the manuscript as follows:

"Surface O_3 concentrations in China are projected to decrease relative to the 2013–2024 annual mean, by 12.6±3.1 $\mu g \cdot m^{-3}$ (±1 SD of the MME), corresponding to a reduction of approximately 16% by 2050, and by 25.3±7.2 $\mu g \cdot m^{-3}$, corresponding to a reduction of approximately 32% by 2100."

(36) Line 492 – Why does SSP245 increase by 2055?

Response: Thank you for the comment. The explanation of the mid-century increase in O_3 under SSP2-4.5 has been clarified in the revised manuscript (Lines 634–638) as follows:

"This mid-century rise is consistent with previous CMIP6 assessments, which show that under SSP2-4.5, precursor emissions decline only gradually while climate warming and photochemical activity increase rapidly before mid-century. As a result, climate-driven ozone enhancement can

temporarily offset emission-driven declines, leading to a short-term increase around 2055 (Gidden et al., 2019; Turnock et al., 2020).”

(37) Line 493 – Is the exceedance of 10 µg m-3 already mentioned earlier?

Response: Thank you for the comment. We have removed the previously mentioned statement about the reduction exceeding 10 $\mu g \cdot m^{-3}$ to avoid redundancy. The revised text in Lines 632–633 is as follows:

“The most significant decrease occurs in South China, where it may reach 24.5 $\mu g \cdot m^{-3}$ by 2100.”

(38) Line 505-506 – Is there not also large differences over South China?

Response: Thank you for the comment. We have clarified that while CMIP6 model differences are most pronounced in Central and East China, there are also smaller differences in Southwest and South China. The revised text in Lines 648–650 now reads:

“Additionally, the projected differences among CMIP6 models are most pronounced over Central and East China, with secondary hotspots of model spread also evident over Southwest and South China, indicating some divergence in the model simulations of O_3 across these regions.”

(39) Line 507-10 - SSP3-7.0-lowNTCF scenario only targets air pollutant controls and nothing to do with carbon controls. Does the lowNTCF scenario not actually increase the contribution to climate change https://acp.copernicus.org/articles/20/9641/2020/?

Response: We thank the reviewer for the comment. The text has been revised Lines 651–655) as follows to clarify that the SSP3-7.0-lowNTCF scenario only implements additional controls on short-lived climate forcers (SLCFs) and does not involve CO_2 or other long-lived greenhouse gas mitigation.

“In the SSP3-7.0-lowNTCF scenario, additional measures targeting non-methane short-lived climate forcers (SLCFs), including aerosols and ozone precursors, are implemented on top of the baseline SSP3-7.0 scenario. These measures are primarily aimed at improving air quality, without additional CO_2 or long-lived greenhouse gas mitigation (Allen et al., 2020). Reductions in SLCFs may slightly affect climate through changes in aerosol radiative effects.”

(40) Line 525-526 – SSP585 actually has strong air pollutant controls https://doi.org/10.1016/j.gloenvcha.2016.05.012

Response: We thank the reviewer for the comment. The text has been revised (Lines 673–675) as

follows to clarify that the SSP5-8.5 scenario features strong air pollutant controls.

"In the SSP5-8.5 scenario, characterized by high radiative forcing, weak climate mitigation, and only limited air pollutant controls (Rao et al., 2017), the annual mean surface O_3 concentration in China is projected to increase by 6.3±1.6 $\mu g \cdot m^{-3}$ by 2050. "

(41) Line 530-531 – do the SSP585 and SSP370-lowNTCF have similar air pollutant emissions and climate change?

Response: We thank the reviewer for the comment. The text has been revised (Lines 677–679) as follows to clarify the observed similarity in surface O_3 changes between SSP5-8.5 and SSP3-7.0-lowNTCF, while noting that the underlying emissions and radiative forcing differ between the two scenarios.

"The projected changes in surface O_3 concentrations for the sub-regions in this scenario are similar to those in the SSP3-7.0-lowNTCF scenario (with a correlation of up to 0.7), despite differences in the underlying emissions and radiative forcing between the scenarios."

(42) Line 539-545 – SSP1-1.9 has strong controls on climate and air pollutants. Careful with the description of effects from this scenario

Response: We thank the reviewer for the comment. The text has been revised (Lines 678–687) as follows to accurately describe SSP1-1.9 as a scenario featuring strong climate mitigation and stringent air pollutant controls.

"Under the Tier 2 experiment, the SSP1-1.9 scenario, characterized by the strongest climate mitigation and the most stringent air pollutant controls, leads to substantial reductions in surface O_3 levels over China. By 2050, the annual mean surface O_3 concentration is projected to decrease by 16.6±7.1 $\mu g \cdot m^{-3}$, and further decline to 25.3±9.5 $\mu g \cdot m^{-3}$ by 2100, corresponding to an overall reduction of approximately 32%. This pronounced decline is primarily driven by the large-scale and coordinated reductions in NO_x and VOCs under this scenario, which markedly suppress the photochemical production of O_3. Meanwhile, although climate-induced changes and concurrent $PM_{2.5}$ reductions can also influence O_3 through complex atmospheric chemical and physical processes, the dominant effect of precursor emission reductions ultimately leads to a substantial overall improvement in surface O_3 air quality under SSP1-1.9."

(43) Line 549-550 – SSP5-3.4-over – this is an overshoot scenarios so will the time evolution be particularly different?

Response: We thank the reviewer for this insightful comment. We have corrected the scenario name throughout the manuscript, replacing all instances of "SSP5-3.4-over" with the official designation "SSP5-3.4-OS", which is an overshoot scenario designed to temporarily exceed its long-term radiative forcing target before declining toward it by 2100. This leads to a temporal evolution that differs from the monotonic forcing pathways in SSP4-3.4 and SSP4-6.0. To clarify this distinction, we have revised the manuscript in Lines 686–711 as follows:

"Under the SSP4-3.4 and SSP4-6.0 scenarios, which are characterized by regionally uneven climate mitigation and air pollutant reductions, the surface O_3 concentration in China is projected to increase by 13.5±1.3 μg·m^{-3} and 18.3±1.4 μg·m^{-3} by 2050, respectively. By 2100, the concentrations under these scenarios are expected to decrease by 9.2±8.3 μg·m^{-3} and 1.6±5.9 μg·m^{-3}, respectively. In contrast, SSP5-3.4-OS is an overshoot scenario, its radiative forcing pathway is designed to temporarily exceed the 3.4 W·m^{-2} level mid-century before declining back to that target by 2100. This pathway involves substantial late-century mitigation efforts, which also lead to reductions in air pollutant emissions as part of the broader deep-decarbonization strategy. Under this scenario, surface O_3 increases by 12.6 ± 1.6 μg·m^{-3} by 2050, comparable to the responses under the moderate-mitigation pathways. However, following the overshoot and the subsequent implementation of stringent mitigation, O_3 concentrations decline markedly by 2100, with a projected reduction of 13.9 ± 11.7 μg·m^{-3}, the largest decrease among the three scenarios.This comparison highlights that although near-term O_3 increases across all three scenarios due to the combined influences of climate change and evolving precursor emissions, sustained and ambitious emission controls, particularly those implemented after an overshoot, as in SSP5-3.4-OS, can ultimately produce substantial long-term reductions in surface O_3, improving air quality and mitigating associated climate impacts."

(44) Line 592-593 – why are NOx emissions so different between the models?

Response: We thank the reviewer for this important comment. The differences in NOx emissions between models arise because each CMIP6 Earth System Model uses its own anthropogenic emission dataset version, model-specific harmonization procedures, and spatial downscaling methods. Even under the same nominal scenario (e.g., SSP3-7.0), models do not ingest a single unified gridded NOx dataset; instead, modeling centers apply scenario-consistent emissions but with differences in sectoral allocation, vertical injection, chemical mechanisms, and pre-processing. These procedural differences can lead to substantial inter-model variations in regional NOx emissions. To clarify this point, we have revised the manuscript as follows (Lines 756–760).

"These inter-model differences in NO_x emissions arise because CMIP6 models use emission inputs derived from different source datasets and processing chains. Each modeling center applies its own downscaling, sectoral allocation, and harmonization procedures to the SSP scenarios, which can lead to notable regional discrepancies even when following the same narrative."

(45) Line 595 – Similar to above why are CH4 concentrations so different given that the input data should be the same?

Response: We thank the reviewer for the comment. For short-lived species such as NO_x, CMIP6 adopts an emission-driven framework. Although all models follow the same SSP narratives, individual modeling centers generate gridded emissions using different underlying inventories, assumptions, and processing methodologies. This leads to inter-model differences in emission inputs, which represent an important source of uncertainty assessed in this study. In contrast, for long-lived greenhouse gases such as CH_4, CMIP6 employs a concentration-driven framework designed to ensure consistent radiative forcing across models. Consequently, inter-model differences in simulated CH_4 concentrations primarily reflect differences in internal chemical processes—particularly OH radical chemistry that controls CH_4 lifetime—rather than differences in prescribed forcing. This has been clarified in the revised manuscript as follows (Lines 760–766).

"In contrast, in MRI-ESM2-0 and EC-Earth3-AerChem, NOx titration is rare during DJF. In addition, EC-Earth3-AerChem exhibits higher simulated CH_4 concentrations (Fig. S4). Although CH_4 forcing is prescribed and harmonized across CMIP6 models, inter-model differences in simulated CH_4 concentrations and their impacts on surface O_3 can still arise from differences in chemical mechanisms, particularly the representation of OH radical fields that control CH_4 lifetime. Higher CH_4 levels in EC-Earth3-AerChem therefore enhance tropospheric background O_3 production, contributing to its relatively higher simulated O_3 concentrations in Central China."

(46) Line 658 – Is this still going to be a VOC-limited regime in the future? Line 660 - Is this still what is happening in the future? Where does the future NOx and VOCs fit on this diagram?

Response: Thank you for the comment. The O_3 formation regime diagnosed in this study is based on present-day observations, and future conditions may differ due to changes in precursor emissions.This point has been clarified in the revised manuscript as follows (Lines 828–841).

"It is important to note that the O_3 formation regime depicted in Figure 15 is derived from present-day observations and should be regarded as a diagnostic reference rather than a fixed projection of future atmospheric chemistry. CMIP6 multi-model simulations under SSP3-7.0 project that NO_x emissions

increase from 2015 and peak around 2035 before gradually declining to near 2015 levels by 2065 (Figure S3), while NMVOCs emissions continue a modest upward trend throughout the century (Figure S6). Based on these projections, our analysis suggests that many regions in China could remain in or near the high-NO_x suppression regime during the first half of the 21st century (around the NO_x emission peak), where O_3 formation could be inhibited by excessive NO_x. As NOx emissions decline thereafter, the chemical environment is projected to gradually shift towards a regime more sensitive to NMVOCs, potentially approaching VOC-limited conditions, where O_3 production becomes increasingly dependent on NMVOCs levels. Regional variations in the relative reduction rates of NO_x and NMVOCs are expected to further influence these transitions. These potential shifts highlight the need for adaptive, region-specific emission-control strategies that account for evolving chemical environments."

Referee #2

Review of "Historical and future changes and present-day uncertainties of ozone in China from CMIP6 models"

The manuscript analyzes historical and future changes, as well as present-day uncertainties, in surface ozone concentrations over China using multiple CMIP6 models and the TAP dataset. It evaluated the performance of nine CMIP6 models and discussed the historical changes of surface ozone over China since 1850. Through comparison of CMIP6 models under a single future scenario, the potential sources of inter-model discrepancies were discussed. Overall, this paper presents a meaningful assessment of present-day surface ozone over China and highlights potential future variations under multiple scenarios. I consider the manuscript suitable for publication after minor revisions addressing the specific comments provided.

1. Specific Comments

(1) The abstract may require some revision. The current version reads more like a conclusion section, and restructuring it to improve clarity.

Response: We thank the reviewer for this helpful suggestion. Following the reviewer's comment, the abstract has been substantially revised.

Revised context are shown below (Lines 16–37):

Ozone (O_3) contributes to global climate change and poses a direct threat to human health. This study investigates the historical variability, future projections, and associated present-day uncertainties of surface O_3 concentrations over China using simulations from nine CMIP6 models and observational data from the Tracking Air Pollution in China (TAP) dataset. A multi-model ensemble mean (MME), constructed using an equal-weighted approach, is used to evaluate model uncertainties across different seasons, underlying surface types, total cloud cover, and $PM_{2.5}$ concentrations, and to assess model performance under future Shared Socioeconomic Pathway (SSP) scenarios. The results show that the MME captures the pronounced seasonal cycle of surface O_3, with higher concentrations during June–August (JJA, ~105 $\mu g \cdot m^{-3}$) and lower values during December–February (DJF, ~55 $\mu g \cdot m^{-3}$), but underestimates O_3 across most regions of China, particularly in East China. Model performance varies with environmental conditions, showing lower bias, MAE, and RMSE over natural land surfaces than over anthropogenic surfaces. The O_3 bias is minimized under cloudy conditions, maximized under partly cloudy conditions, and generally increases with rising $PM_{2.5}$ concentrations before declining beyond a threshold. Over the historical period, the MME simulates a substantial increase in annual mean surface O_3 across China (~39.3 $\mu g \cdot m^{-3}$). Future projections indicate continued O_3 increases under weak mitigation (SSP3-7.0), with East China rising by 26.9%, and widespread decreases under strong mitigation (SSP1-2.6), particularly in Southwest and South China (>30 $\mu g \cdot m^{-3}$). Analysis of model

spread and its drivers indicates that uncertainties in surface O_3 projections arise from the combined effects of emissions (including precursors and $PM_{2.5}$), climate conditions, and model representations of chemistry and circulation. Improving the understanding of these coupled influences is essential for enhancing the reliability of regional O_3 projections and for informing effective air quality and climate mitigation strategies in China.

(2) It would be helpful to provide additional information on the source and basic characteristics of the cloud cover data used in Section 3.3.

Response: Thank you for this helpful suggestion. To clarify the source and basic characteristics of the cloud cover data used in Section 3.3, we have added a new supplementary table (Table S1: CMIP6 variables and auxiliary datasets used in this study). This table now explicitly documents the data source, temporal resolution, spatial resolution, and purpose in the analysis for the cloud cover variable, thereby improving transparency and reproducibility.

Revised Tables are shown below:

Table S1 – CMIP6 variables and auxiliary datasets used in this study.

Variable name	Source	Temporal resolution	Purpose in the analysis	Reference
Ozone	CMIP6 models	Amon	O_3 analysis	Eyring et al. (2016)
	TAP dataset	Mon	Uncertainties analysis	Xue et al. (2020)
	Ground-based observations	Mon (derived from hourly)	O_3-NO_2-NMVOCs relationship	Li et al. (2017)
NO_x	CMIP6 models	Amon	Correlation analysis	Collins et al. (2017)
	Ground-based observations	Mon (derived from hourly)	O_3-NO_2-NMVOCs relationship	Li et al. (2017)
TAS	CMIP6 models	Amon	Correlation analysis	Eyring et al. (2016)
CH_4	CMIP6 models	Amon	Correlation analysis	Collins et al. (2017)
NMVOCs (emivoc)	CMIP6 models	Amon	Correlation analysis	Collins et al. (2017)
BVOCs (emibvoc)	CMIP6 models	Amon	Correlation analysis	Collins et al. (2017)
Surface types	MODIS/MCD12Q1	2020 (fixed)	Uncertainties analysis	Friedl and Sulla-Menashe (2019)
Total cloud cover	ERA5 (ECMWF)	Mon	Uncertainties analysis	Wu et al. (2023)
$PM_{2.5}$ and its components	TAP dataset	Mon	Uncertainties analysis	Xiao et al. (2022)

(3) Lines 147–151: The manuscript states that "due to the limited availability of model data for the Tier 2 CMIP6 scenarios (SSP1-1.9, SSP4-3.4, SSP4-6.0, and SSP5-3.4-over), the analysis focuses on SSP3-7.0 and the Tier 1 scenarios." However, it is unclear which scenarios are included in the MME used in the analysis. Additionally, it is not evident where the Tier 2 scenarios were applied in the study. I recommend clarifying these points in Section 2, specifying exactly which SSP scenarios are used in the MME and explaining whether/where the Tier 2 scenarios contribute to the analysis.

Response: We thank the reviewer for this insightful comment. In this study, all core quantitative analyses and the multi-model ensemble (MME)—including present-day uncertainty assessment, historical evaluation, and inter-model spread analysis—are conducted exclusively using Tier 1 CMIP6 scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP3-7.0-lowNTCF, and SSP5-8.5), which provide the most extensive and internally consistent multi-model coverage. Tier 2 scenarios (SSP1-1.9, SSP4-3.4, SSP4-6.0, and SSP5-3.4-OS) are used only in Section 5.1 to qualitatively illustrate potential future surface O_3 trajectories under alternative emission pathways.

We have revised Section 2 (Lines 165–174) to explicitly specify the SSP scenarios included in the MME and to clarify the limited and contextual role of the Tier 2 scenarios. In addition, a new Table S4 has been added to the Supplementary Material to summarize the meaning and key characteristics of the typical CMIP6 scenarios.

"Due to the limited availability and inconsistent model coverage of Tier 2 CMIP6 scenarios (SSP1-1.9, SSP4-3.4, SSP4-6.0, and SSP5-3.4-OS), all present-day uncertainty analyses and multi-model spread analysis in this study are conducted exclusively using Tier 1 scenarios. Tier 2 scenarios are included in Section 5.1 only to illustrate potential future trajectories of surface O_3 under alternative emission pathways. Some numerical changes are reported to highlight relative differences between these scenarios, but they are not used for robust multi-model inter-scenario comparisons or quantitative uncertainty assessment. Although one CMIP6 model (MRI-ESM2-0) provides simulations for all Tier 1 and Tier 2 scenarios, single-model inter-scenario comparisons are not performed, as our analysis relies on multi-model ensemble statistics to ensure representative and robust conclusions."

Table S4 – Meaning and characteristics of typical scenarios in CMIP6 in this study

（Global warming levels are based on the IPCC AR6 assessments of CMIP6 model projections (2081–2100 vs. 1850–1900) with likely (5–95 %) ranges. Scenarios marked * are research variants and not part of the five core SSP scenarios.）

Scenario	Socioeconomic Pathway	2100 Radiative Forcing ($W \cdot m^{-2}$)	Global Warming in 2100 (°C, relative to 1850–1900)	Emission Trends & Policy Features	Typical Applications / Notes	Reference
SSP1-1.9	Sustainable development (green growth, low inequality, high tech)	≈1.9	~1.5°C (likely range: 1.2–1.7°C)	Strong mitigation; net-zero before 2050	1.5°C pathway for Paris Agreement evaluation	Riahi et al. (2017)
SSP1-2.6	Sustainable development (same as SSP1)	≈2.6	~1.8°C (likely range: 1.37–2.4°C)	Clean energy transition; moderate mitigation	Low-emission sustainable development scenario	IPCC AR6 WGI (2021)
SSP2-4.5	Medium pathway (current trends, moderate mitigation)	≈4.5	~2.7°C (likely range: 2.1–3.5°C)	Limited global cooperation; gradual emission slowdown	Baseline “medium emission” scenario	Riahi et al. (2017)
SSP3-7.0	Regional rivalry (fragmented development, low cooperation)	≈7.0	~3.6°C (likely range: 2.8–4.6°C)	Developing countries grow rapidly; fossil fuels dominate	Used as pessimistic reference scenario	IPCC AR6 WGI (2021)
SSP3-7.0-lowNTCF*	Same as SSP3-7.0 but with NTCF control	≈7.0	Not officially assessed in AR6	Stronger control of CH_4, NMVOCs, BC	Research variant; not an IPCC core scenario; temperature response varies by model	Lund et al. (2020)
SSP4-3.4	Inequality pathway (widening North–South gap)	≈3.4	~2.3°C (likely range: 1.9–2.8°C)	Developed countries reduce emissions; developing nations increase	Medium-low emission; studies of inequality impacts	O’Neill et al. (2017)
SSP4-6.0	Same as SSP4 but with higher global emissions	≈6.0	~3.0°C (likely range: 2.8–4.0°C)	Strong disparity between high- and low-emission nations	Medium-high emission; reflects global inequality	O’Neill et al. (2017)
SSP5-8.5	Fossil-fueled growth (high growth, no mitigation)	≈8.5	~4.4°C (likely range: 3.8–5.7°C)	Very high emissions; no climate policies	Extreme high-emission “worst case” scenario	Riahi et al. (2017)
SSP5-3.4-OS*	Fossil-fueled development (tech-optimistic); overshoot pathway	≈3.4 (overshoot)	Not officially assessed in AR6	High early emissions; large-scale carbon removal later	Overshoot research scenario; peak warming >2°C then declines; not AR6 official estimate	O’Neill et al. (2016)

(4) Lines 165–179: The necessity of analyzing TAP data in relation to PM2.5 and its chemical components is unclear. Please clarify. If this analysis does not directly support the study’s objectives or contribute to the main conclusions, I suggest removing this part to maintain focus and coherence.

Response: We thank the reviewer for this constructive comment. We agree that the analysis of relationships between TAP surface O_3 and $PM_{2.5}$ chemical components is not central to the main objectives of this study, which focus on CMIP6-based evaluation and projection of surface O_3. To enhance the focus and coherence of the manuscript, we have removed this analysis from the main text. Any relevant points have been either omitted or briefly summarized to support the interpretation of results where necessary. The manuscript has been revised accordingly.

(5) In Section 3.1 (Lines 218–224), the manuscript would benefit from a clearer explanation of why surface ozone concentrations remain high across multiple seasons in Northwest China, specifically whether the elevated levels are driven primarily by local photochemical production or by regional transport? In addition, it would be helpful to elaborate on the reasons for the comparatively lower ozone levels observed in Northeast China.

Response: Thank you for the valuable comment. In Northwest China, surface O_3 concentrations are high throughout the year, with a seasonal cycle that peaks in JJA. The persistently elevated O_3 levels result from multiple factors, We have added an explanation in the revised manuscript (Lines 264–278) as follow:

"In Northwest China, O_3 concentrations remain high across all four seasons, although a seasonal cycle is present with the highest values in JJA (113 $\mu g \cdot m^{-3}$).The high O_3 levels can be attributed to multiple factors. First, the region is arid and receives strong solar radiation, which promotes efficient photochemical O_3 production. Second, although overall anthropogenic emissions are relatively low compared with eastern and southern China, some cities such as Urumqi and Yinchuan still exhibit relatively high O_3 concentrations during summer, indicating that local emissions from industrial, energy, and traffic activities cannot be neglected (Zhu et al., 2023). Third, Northwest China can be affected by long-range transport of O_3 and its precursors from surrounding and remote source regions. Modeling studies have shown that European anthropogenic emissions can enhance surface O_3 over northwestern China by approximately 2–6 ppbv during spring and summer, while contributions from the Middle East may occasionally add another 1–4 ppbv (Li et al., 2014). Finally, high-altitude areas are more susceptible to STE, allowing O_3 from the upper troposphere and lower stratosphere to be transported downward, further elevating surface O_3 background levels (Yin et al., 2023)."

Meanwhile, we have added an explanation for the relatively low O_3 concentrations in northeast China in the revised manuscript (Lines 278–286) as follows:

"Northeast China, on the other hand, has the lowest O_3 concentrations among all sub-regions, particularly during MAM, JJA, and SON, with concentrations of only 76 $\mu g \cdot m^{-3}$, 88 $\mu g \cdot m^{-3}$, and 54 $\mu g \cdot m^{-3}$, respectively. This can be attributed to several factors: first, the cooler climate and shorter period of strong solar radiation in northeast China slow down the rate of photochemical ozone production. Second, the region has relatively lower emissions of reactive VOCs compared to the more industrialized and densely populated regions of eastern and southern China, which constrains the overall ozone production potential. Third, northeast China is frequently influenced by clean continental air masses from Siberia, which effectively dilute local and regional pollutant concentrations."

(6) In Section 3.1, why UKESM1-0-LL exhibits the largest underestimation? A discussion of potential reasons would aid in interpreting the model results.

Response: We thank the reviewer for this insightful comment. We have revised the manuscript to provide a brief discussion of why UKESM1-0-LL exhibits the largest underestimation of surface O_3 (Lines 322–333).

The specific modifications are as follows:

"The nine CMIP6 models evaluated in this study generally underestimate O_3 concentrations across most sub-regions of China. The most pronounced underestimations occur in UKESM1-0-LL and UKESM1-1-LL. Although these two models show slight overestimations in Southwest China from June to September, as well as in South China in June and in Northwest China in August, underestimation remains the dominant pattern, particularly in East and Central China during MAM and DJF, where simulated O_3 concentrations are over 60 $\mu g \cdot m^{-3}$ lower than the TAP values. This is consistent with the findings of Turnock et al. (2020), who suggested that these biases may be related to excessive NO_x titration in the UKESM1-0-LL model, leading to an underestimation of surface O_3 concentrations over much of the Northern Hemisphere continental regions during DJF. Liu et al. (2022) further conducted a global evaluation of surface O_3 biases in UKESM1, identifying temperature, photolysis rates, and regional chemical composition as key predictors, which may help explain the underestimations identified in this study. "

(7) In Section 3.4, could the manuscript provide a more detailed discussion of the primary mechanisms through which aerosols affect surface ozone?

Response: Thank you for the comment. We have revised the text to specify how aerosols affect O_3 formation. The changes can be found in Lines 493–499 as follows:

"Aerosols play a crucial role in the simulation of surface O_3 concentrations. Variations in $PM_{2.5}$ can influence O_3 formation through several pathways: (1) chemical effects: changes in aerosol composition (e.g., sulfate, nitrate, organic matter, BC) can perturb the concentrations of key radicals (e.g., OH, HO_2) and thus alter photochemical reaction rates (Wang et al., 2019); (2) radiative effects: aerosols modify optical properties and reduce actinic flux at the surface, thereby suppressing photolysis rates and slowing O_3 production; and (3) heterogeneous effects: reactions on aerosol surfaces can directly consume NO_x or other precursors, shifting the chemical environment for O_3 formation (Qu et al., 2021)."

(8) Section 5. Splitting it into several sub-sections could help improve the clarity of the manuscript.

Response: Thank you very much for this helpful suggestion. We agree that Section 5 was too long in the original manuscript. To improve readability, we have now divided Section 5 into several sub-sections:

5.1 Projected trends under different SSP scenarios

5.2 Model spread and its drivers under the SSP3-7.0 scenario

These structural adjustments make the section easier to follow.

(9) Section 6 (Summary) could be made more concise.

Response: Thank you for your valuable comments. In response to your suggestions, we have thoroughly revised the section to significantly enhance its synthesis and discussion depth.

We have clarified this point in the revised manuscript Lines 920–968 as follows:

"This study provides a comprehensive evaluation of surface O_3 over China using the CMIP6 multi-model ensemble, spanning present-day performance, historical evolution (1850–2014), and future projections under diverse scenarios. By synthesizing findings across these temporal scales, we offer an integrated perspective on model capabilities, key sources of uncertainty, and the implications for understanding past changes and future air quality. The principal insights are discussed below.

(1) The CMIP6 MME reproduces the observed spatial and seasonal pattern of surface O_3, with maxima in summer and over central China. However, a systematic underestimation prevails across most models and regions, particularly in eastern China. This bias is not random; rather, it tends to be modulated by environmental conditions: model error generally increases over urban surfaces and under high $PM_{2.5}$ concentrations, revealing a complex, non-linear relationship especially in winter. Critically, these identified present-day biases are instructive for interpreting simulations of other periods. They point to persistent structural model limitations in key processes, particularly photochemical mechanisms and aerosol-chemistry interactions. Consequently, historical or future periods characterized by environmental conditions analogous to those linked with high bias today are likely subject to similar directional biases in simulated O_3. Acknowledging and accounting for this transferable understanding of model bias can therefore help advance a more robust interpretation of both past changes and future projections.

(2) The MME indicates a substantial increase of 39.3 ± 14.4 $\mu g \cdot m^{-3}$ in China's annual mean surface O_3 since 1850, with a marked acceleration post 1950. The large inter-model spread in this trend originates primarily from two factors highlighted in the literature: (i) differences in the simulated pre-industrial

(1850) baseline concentrations, and (ii) divergent chemical sensitivities of O_3 production to emission changes. This spectrum of sensitivity is clearly evident in our results: MIROC-ES2H exhibits high sensitivity and simulates the largest historical trend, whereas UKESM1-0-LL shows lower sensitivity, producing a trend most consistent with a direct, emission-driven response. The fact that the greatest model disparity occurs in Central China, a region also undergoing significant aerosol changes, further highlights the contribution of aerosol-chemistry interactions to the overall projection uncertainty.

(3) Future surface O_3 levels over China are governed by the complex interplay between air pollution control and climate mitigation. A pronounced decrease occurs under the strong co-mitigation scenario SSP1-2.6. In stark contrast, the high-forcing SSP3-7.0 pathway projects continuous O_3 increases in many regions. Our analysis reveals that the common projection of a mid-century O_3 peak followed by a decline can arise from two fundamentally different policy dominances: pathways where climate mitigation objectives are central (SSP2-4.5), and pathways where improvements are primarily driven by stringent air quality controls even under high climate forcing (SSP5-8.5). This underscores a critical insight: substantial O_3 reductions are achievable under divergent socioeconomic futures, but they demand either deep coupled climate-air quality governance or, alternatively, exceptionally strong dedicated air pollution policies in the absence of climate action.

(4) This uncertainty arises from fundamental differences in how models represent key chemical and climatic processes. Under SSP3-7.0, while models generally agree on the positive roles of climate warming (TAS), global background chemistry (CH_4), and precursors (NMVOCs), they diverge sharply in simulating the local efficacy of NO_x control, as indicated by inconsistent O_3-NO_x correlations. Such divergence underscores that the simulated O_3 response is not merely a function of emission inputs but is critically dependent on each model's representation of complex, non-linear transition between chemical regimes (e.g., VOC-limited vs. NO_x-saturated), which can be further modulated by model-specific climate sensitivities and background emission levels. Consequently, the optimal strategy for future O_3 abatement, specifically, the relative priority and expected benefit of controlling NO_x versus managing other drivers like CH_4 or NMVOCs under a warming climate, remains highly model-dependent. This poses a fundamental challenge for deriving robust, model-consistent policy insights, highlighting the need to interpret projections within the context of these underlying chemical mechanisms.

2. Comments on figures and tables

(1) Figure 1, the $PM_{2.5}$ components are difficult to discern. Since the focus of this study is on ozone, it is recommended to either remove this panel or adjust it to improve clarity.

Response: We thank the reviewer for this comment. We agree that the original Figure 1 was visually

cluttered, making the aerosol components and text hard to discern, and it did not directly support the ozone-focused narrative. To address this, we have replaced the original figure with a new, clear map illustrating the geographic locations of China and its seven sub-regions used throughout the study. This revision serves two purposes: (1) it eliminates the legibility issues by presenting essential spatial information in a clean, readable format; and (2) it provides a foundational reference for the regional analyses that follow, thereby strengthening the manuscript′s focus.

Revised Figure 1 is shown below:

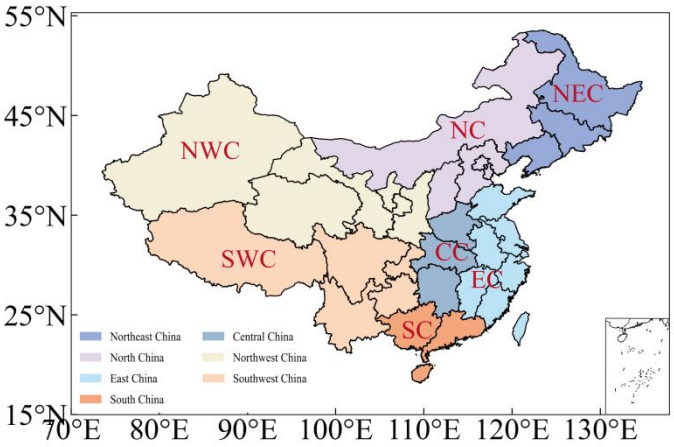

Figure 1. Geographic locations of China and its seven sub-regions used in this study (Northeast China (NEC), North China (NC), East China (EC), South China (SC), Central China (CC), Northwest China (NWC), and Southwest China (SWC)).

(2) Figure 2, the Bias color scheme may be misleading, as the current gradient mixes small negative and positive values. A more effective scheme would assign white to zero, blue gradients to negative values, and red gradients to positive values, enhancing clarity.

Response: Thank you for this helpful suggestion. We agree that a zero-centred colour scheme is essential for clearly distinguishing negative and positive Bias values. In the revised manuscript, we have updated Figure 2 by adopting a diverging colour map for Bias, with white representing zero, blue shades indicating negative values, and red shades indicating positive values.

In addition, we now apply separate colour bars for each column according to the specific metric being displayed, rather than using a single shared scale. This combined revision significantly improves visual clarity and avoids potential misinterpretation caused by mixing different statistical metrics on the same colour scale.

The revised figure is shown below.

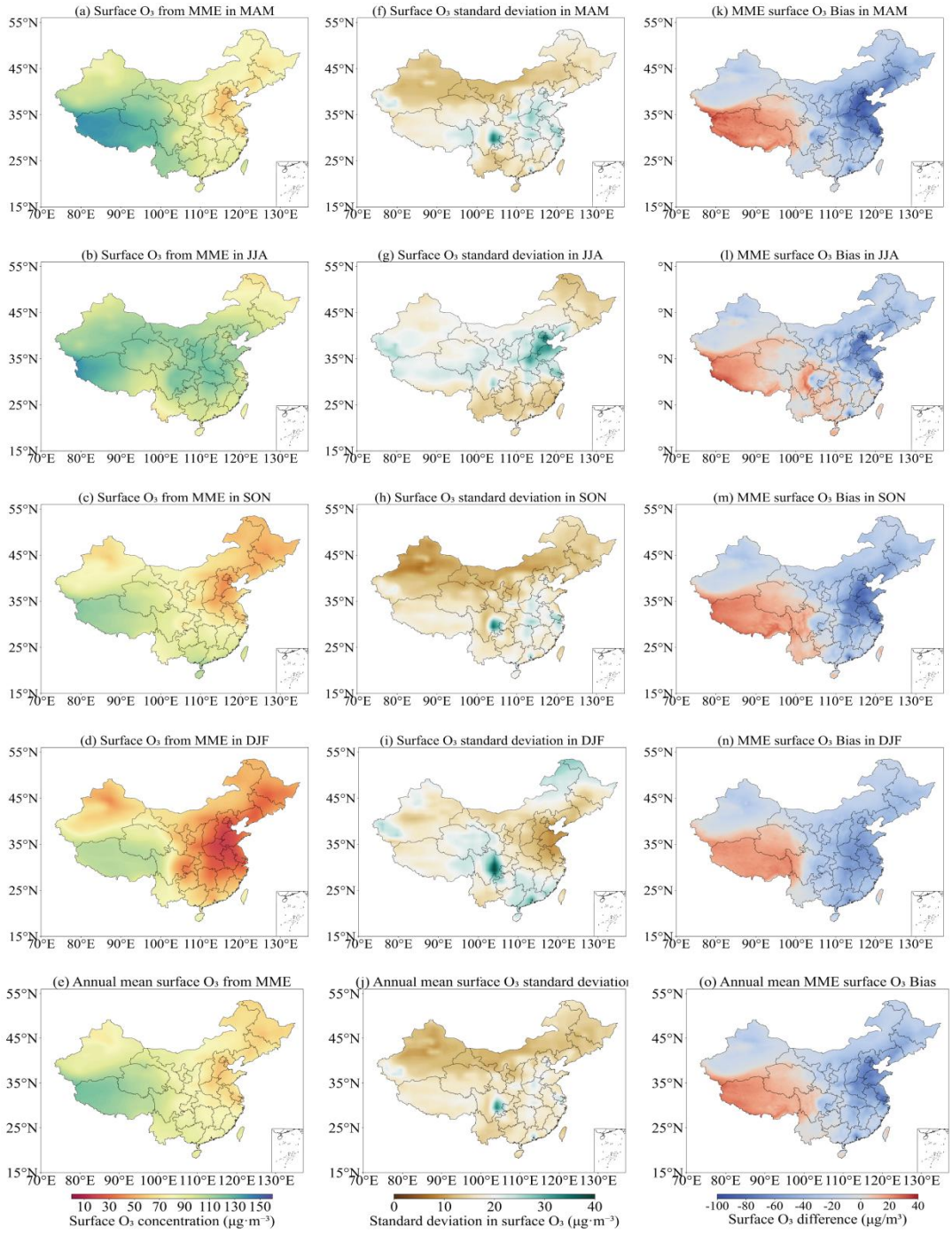

Figure 2. Multi-model (nine CMIP6 models) annual and seasonal mean surface O_3 concentrations over the 2015–2023 period in (a) MAM; (b) JJA; (c) SON; (d) DJF; and (e) annual mean. The SD of the MME in (f)MAM, (g) DJF, (h) JJA, (i)SON, and (j) annual mean. The difference between the MME and TAP observations in (k)MAM, (l) DJF, (m) JJA, (n)SON, and (o) annual mean.

(3) Figures 3 and 10. Adjusting the color or line style would improve differentiation from individual CMIP6 models and enhance figure clarity.

Response: Thank you for the helpful suggestion. We agree that in the original Figure 3 and 10 the

MME curve was not sufficiently distinguishable, which may have limited the clarity of the comparison with TAP. In the revised Figure 3 and 10, we have made the following improvements to enhance readability: (1) increased the line thickness and saturation of the MME curve, and (2) slightly reduced the opacity of the TAP shading. These adjustments improve the visual contrast and make the MME-TAP comparison clearer across all regions.

Revised Figure are shown below:

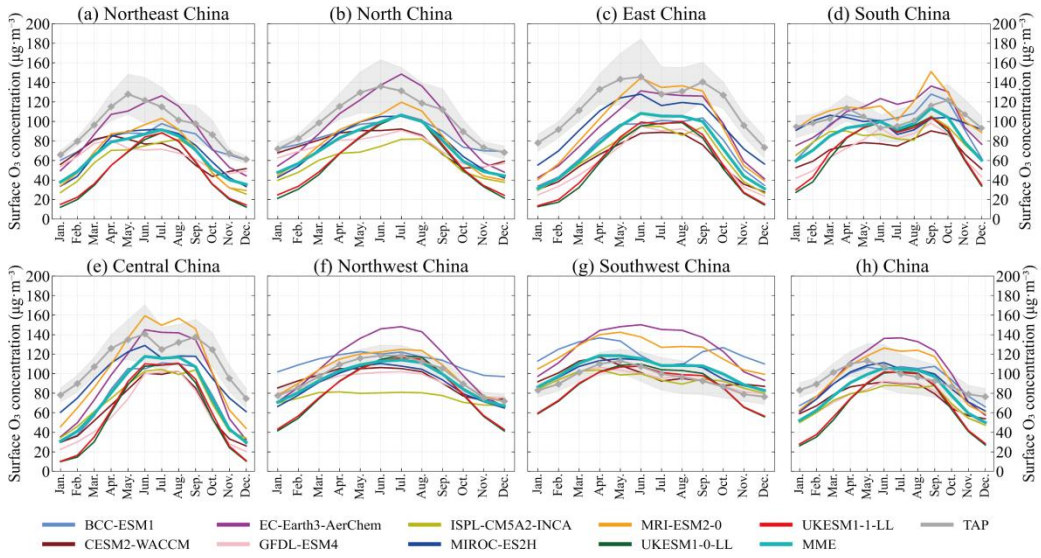

Figure 3. Comparison of the annual cycle of O_3 concentrations, between individual CMIP6 models, the MME and TAP in China and sub-regions for the period 2015–2023. The grey shading shows SD of TAP observations within the region

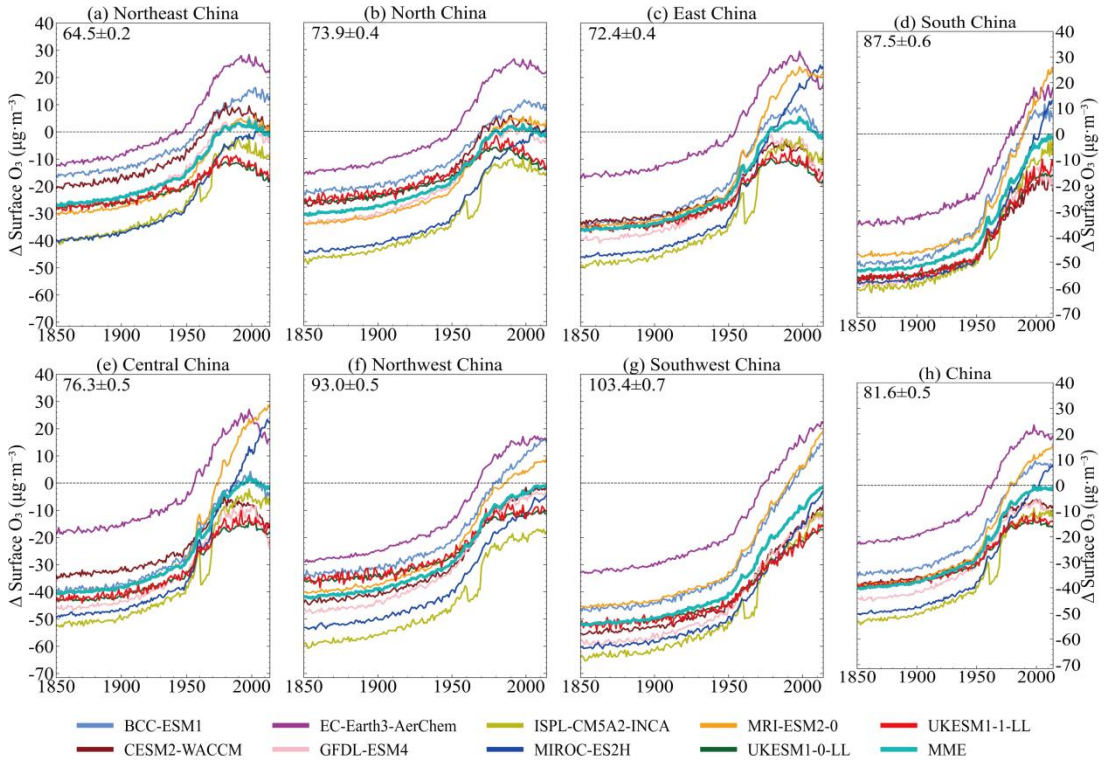

Figure 10. Changes in the China and sub-regions annual mean surface O_3 concentrations from the pre-industrial period to present day, relative to a 2015–2023 mean value, across nine CMIP6 models and MME. The multi-model annual mean 2015 – 2023 surface O_3 concentrations (±1 SD) are shown in the top left of each panel.

(4) It would be helpful to include a summary table in the supplementary materials presenting key statistical metrics for each model. This would allow readers to more clearly compare model performance and improve the clarity of the results.

Response: We thank the reviewer for the suggestion. In response, we have added the relevant statistics in the Supplementary Materials: Table S2 summarizes the seasonal performance statistics of the multi-model ensemble (MME) O_3 relative to TAP, and Table S3 provides the annual correlation coefficients (r) between individual CMIP6 models and TAP.

Revised Tables are shown below:

Table S2. Summary of seasonal performance statistics of MME O_3 relative to TAP

Statistics (μg·m^{-3})	MAM	JJA	SON	DJF
Bias	-25.4	-12.8	-22.9	-25.3
SD (Inter-model)	16.8	18.8	15.2	19.0

Table S3. Annual correlation coefficients (r) between individual CMIP6 models and TAP.

Models	BCC-ESM1	CESM2-WACCM	EC-Earth3-AerChem	GFDL-ESM4	IPSL-CM5A2-INCA	MIROC-ES2H	MRI-ESM2-0	UKESM1-0-LL	UKESM1-1-LL	MME
r	0.78	0.91	0.77	0.80	0.80	0.88	0.81	0.73	0.74	0.80

(5) It is recommended to include basic information on the Tier 2 CMIP6 scenarios in Table 1, or alternatively provide an additional table in the Supplementary Material summarizing the Tier 2 scenarios (e.g., data availability and participating models). This would help readers clearly distinguish between Tier 1 and Tier 2 scenarios and better understand the data selection in this study.

Response: We thank the reviewer for this helpful suggestion. To improve clarity regarding the use of Tier 2 CMIP6 scenarios, we have revised Table 1 to include the Tier 2 scenarios that are used in this study, together with their corresponding data availability and participating models.

This revision allows readers to clearly distinguish between Tier 1 and Tier 2 scenarios and better understand the data selection strategy adopted in our analysis.

Revised Tables are shown below:

Table1. Number of ensemble members used for the historical- and future-scenario experiments from each model in the analysis of surface O_3 in this study.

CMIP6 Models	Institution	Resolution	Historical	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP3-7.0-lowNTCF	SSP5-8.5	SSP1-1.9	SSP4-3.4	SSP4-6.0	SSP5-3.4-OS	Model reference
BCC-ESM1	Beijing Climate Center, China Meteorological Administration, China	2.813° ×2.813°	3			3	3						Wu et al. (2020)
CESM2-WACCM	National Center for Atmospheric Research, Climate and Global Dynamics Laboratory, USA	1.25° ×0.94°	3			3	3						Emmons et al. (2020)
EC-Earth3-AerChem	European Consortium of Meteorological Services, Research Institutes, and High-performance Computing Centers	3.0° ×2.0°	4		1	3	3						Noije et al. (2021)
GFDL-ESM4	NOAA Geophysical Fluid Dynamics Laboratory, USA	1.0° ×1.25°	1	1	1	1	1	1	1				Horowitz et al. (2020)
ISPL-CM5A2-INCA	Institute Pierre Simon Laplace, Paris, France	3.75° ×1.875°	1			1	1						Sepulchre et al. (2020)
MIROC-ES2H	University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine‐Earth Science and Technology, Japan	2.813° ×2.813°	3	1	2	1		3	1				Hajima et al. (2020)
MRI-ESM2-0	Meteorological Research Institute, Japan	2.813° ×2.813°	10	4	10	5	3	5	5	1	1	1	Yukimoto et al. (2019)
UKESM1-0-LL	Natural Environment Research Council, and Met Office, United Kingdom	1.875° ×1.25°	18	5	5	19	1	4					Sellar et al. (2019)
UKESM1-1-LL	Natural Environment Research Council, and Met Office, United Kingdom	1.875° ×1.25°	1	3		1							Mulcahy et al. (2023)
Total number of models			44	14	19	37	15	13	7	1	1	1	