Reviewer: 1
Comments to the Author:
This manuscript compares global-scale simulation results of different ozone stress parameterization schemes in CLM5. The research topic is of strong scientific significance and value for model development. Using consistent meteorological and ozone forcing data, the study conducts a mechanism-consistent evaluation of three representative schemes—Sitch, Lombardozzi, and Li—and performs multi-scale validation against GPP data from MODIS and FLUXNET. Overall, the research framework is relatively comprehensive, and the results are reasonably convincing. In particular, the analysis of how ozone flux thresholds, response function forms, and memory effects influence simulation outcomes has practical significance for improving ozone stress processes in land surface models. However, the manuscript still has room for improvement in terms of the following aspects. The authors are therefore encouraged to revise the manuscript further to enhance its completeness and persuasiveness.

Response: Thank you very much for your positive evaluation of our manuscript and constructive suggestions. We greatly appreciate your recognition of the scientific significance of this work, particularly regarding the comparison of ozone stress parameterization schemes and their implications for land-surface model development. Following your comments, we have carefully revised the manuscript to further improve the clarity, completeness, and robustness of the study. Specifically, we expanded the Introduction to better clarify the scientific rationale and significance of ozone-stress parameterization schemes, improved the interpretation of the no-ozone baseline experiment, and added further discussion regarding the influence of model structural differences (e.g., CLM4.5 versus CLM5) on ozone-stress evaluation. We also refined the descriptions of the results, discussion, and conclusions to improve logical consistency and scientific interpretation across multiple spatial and temporal scales. All detailed revisions corresponding to your comments are provided in the point-by-point responses. In the revised manuscript, modifications made in response to your comments are highlighted in **blue**, while revisions associated with another reviewer are highlighted in **red** for clarity. We sincerely appreciate your valuable comments and guidance, which have substantially improved the quality, rigor, and readability of the manuscript.

Specific comments:
1. It is recommended to further clarify the significance and scientific basis of adopting ozone stress parameterization schemes.

Response: Thank you very much for this important suggestion. Following your comment, we further clarified the significance and scientific basis of adopting ozone-stress parameterization schemes in the Introduction. Specifically, we added an explanation emphasizing that ozone damage is fundamentally controlled not only by ambient ozone concentration, but also by stomatal uptake, plant detoxification/tolerance capacity, and the subsequent physiological regulation of photosynthesis and stomatal conductance. We also clarified why process-based ozone-stress parameterizations are necessary to link atmospheric ozone forcing terrestrial carbon-cycle responses in land surface models.

The following text has been added to the Introduction:

In land surface models, the scientific rationale for adopting ozone-stress parameterizations is that ozone injury is not directly determined by ambient concentration alone, but by the amount of ozone entering leaves through stomata, the capacity of plants to detoxify or tolerate that flux, and the subsequent physiological downregulation of photosynthesis and stomatal conductance (Sitch et al., 2007). Without such process-based parameterizations, models either neglect an important atmospheric stressor or treat ozone exposure in a way that is disconnected from canopy physiology, water limitation, phenology, and differences in plant functional type (Fuhrer, 1996). Therefore, ozone-stress parameterizations provide the necessary mechanistic bridge between atmospheric ozone forcing and terrestrial carbon-cycle responses, allowing land surface models to diagnose when ozone damage is triggered, how it accumulates over time, and how strongly it suppresses GPP across different ecosystems.

2. The current analysis mainly focuses on differences among the various ozone stress parameterization schemes. However, the results suggest that simulations without ozone stress sometimes exhibit higher accuracy. Therefore, it is recommended to include a comparative analysis between results “with ozone stress” and “without ozone stress.”

Response: Thank you very much for this insightful suggestion. Following your recommendation, we substantially expanded the comparison and discussion between simulations “with ozone stress” and “without ozone stress” throughout the manuscript. In the revised Results section, we added a series of explicit comparisons between the no-ozone experiment and the ozone-stress parameterization schemes across multiple spatial, temporal, and biome scales. Specifically, we clarified that the no-ozone case substantially overestimates tropical GPP relative to MODIS and FLUXNET, particularly in humid low-latitude ecosystems and central African regions, underscoring the importance of accounting for ozone stress to reduce excessive tropical productivity. At the same time, we also emphasized that the no-ozone case sometimes exhibits lower RMSE, higher NSE, or stronger preserved seasonal variability than the ozone-stress parameterization schemes in some temperate, boreal, and forest-dominated regions, suggesting that introducing ozone stress alone cannot fully resolve all discrepancies relative to observations and that other physiological or environmental constraints may still contribute to model uncertainty.

To further interpret these findings, we added a new discussion in Section 4 comparing our CLM5-based results with earlier CLM4.5-based studies. We clarified that the baseline no-ozone global GPP in CLM5 (~132 Pg C yr^{-1}) is already substantially lower than that reported in earlier CLM4.5 studies (~146.5 Pg C yr^{-1}), indicating that improvements in land-surface physiology and canopy-process representation have already reduced part of the intrinsic productivity bias. Consequently, the remaining correction space for ozone stress in CLM5 is smaller, and excessive ozone suppression can more easily lead to underestimation in some regions and statistical metrics. We therefore emphasized that the apparent benefit of ozone-stress parameterizations strongly depends on the baseline behavior of the host land-surface model.

In addition, we incorporated this interpretation into the Conclusion section, where we further highlighted that ozone stress effectively suppresses the tropical overestimation in the no-ozone experiment, but that improvements in model performance remain spatially and temporally heterogeneous across different metrics and ecosystems. We also stressed that future ozone parameterizations should aim not only to represent ozone damage itself, but also to achieve balanced interactions with canopy processes, phenology, hydrological constraints, and biome-dependent productivity dynamics.

The following text has been added to the Results:

Compared with the MODIS observations (Fig. 1f), the no-ozone case (Fig. 1a) shows pronounced overestimation of GPP in tropical regions, particularly in humid low-latitude ecosystems, indicating that neglecting ozone stress can substantially exaggerate tropical productivity. Compared with FLUXNET, the no-ozone case (Fig. 2a) exhibits relatively large RMSE values in central African regions, reflecting pronounced overestimation errors under humid tropical conditions. However, at some stations in the middle and high latitudes, the no ozone case still shows lower RMSE values than the ozone-stress parameterization schemes. The no-ozone case generally exhibits smaller biases than the ozone-stress parameterization schemes in several temperate and boreal biomes, suggesting that introducing ozone stress alone does not necessarily improve model performance and that additional physiological or environmental constraints may still be required to better reproduce observations.

The no-ozone case generally produces larger seasonal amplitudes than the ozone-stress parameterization schemes and, in several latitude bands, remains closer to FLUXNET during peak growing seasons, suggesting that introducing ozone stress alone cannot fully resolve the discrepancies between simulations and observations. In these forest PFTs, the no-ozone case generally produces larger seasonal amplitudes and remains closer to FLUXNET during peak growing periods than most ozone-stress parameterization schemes do. Compared with ozone-stress parameterization schemes, the no-ozone case generally exhibits higher normalized standard deviations across several PFTs, indicating stronger preservation of seasonal variability, although its correlations are not consistently the highest, and its performance for shrubs and crops is poorer than that of some ozone-stress parameterization schemes.

Among the model experiments, the no-ozone case generally maintains the highest NSE values in spring, summer, and autumn, indicating that the baseline simulation without ozone stress still reproduces seasonal variability better than most ozone-stress parameterization schemes at large scales. The no-ozone case consistently shows lower RMSE values than all ozone-stress parameterization schemes at the annual, seasonal, and monthly scales, suggesting that additional sources of model uncertainty beyond ozone stress may still contribute to discrepancies relative to FLUXNET. The no-ozone case reaches 132.00 Pg C yr－1, substantially exceeding both MODIS and FLUXCOM estimates, indicating that neglecting ozone stress likely leads to an overestimation of global terrestrial productivity.

The following text has been added to the Discussion:

Compared with the earlier CLM4.5-based study by Lombardozzi et al. (2015), an important difference emerges in the baseline productivity simulated in the absence of ozone stress. In CLM4.5, the no-ozone global GPP reached approximately 146.5 Pg C yr^{-1}, whereas the corresponding no-ozone simulation in the present CLM5 framework is substantially lower (132 Pg C yr^{-1}). This indicates that improvements in the underlying land-surface physiology and canopy process representation from CLM4.5 to CLM5 have already reduced part of the excessive global productivity bias. Consequently, the introduction of ozone stress in CLM4.5 moved the simulated GPP closer to observational benchmarks at the global scale, whereas in CLM5, the remaining correction space is smaller, and excessive ozone suppression can more easily lead to underestimation in some regions and statistical metrics. This difference highlights that the apparent performance of ozone-stress parameterizations is strongly dependent on the baseline behavior of the host land-surface model.

The following text has been added to the Conclusion:

Our results further suggest that the apparent benefit of introducing ozone stress depends strongly on the baseline behavior of the host land-surface model. Compared with earlier CLM4.5-based studies, the no-ozone simulation in CLM5 already produces substantially lower global GPP, indicating that improvements in canopy physiology and land-surface process representation have reduced part of the intrinsic productivity bias. Consequently, although ozone stress effectively suppresses the tropical overestimation present in the no-ozone experiment, improvements in model performance remain spatially and temporally heterogeneous across different evaluation metrics and ecosystems. This highlights that future ozone parameterizations should not only focus on representing ozone damage itself, but also on achieving balanced interactions with the underlying model structure, including canopy processes, phenology, hydrological constraints, and biome-dependent productivity dynamics.

The zonal mean results are currently presented primarily as line plots (Figure 3), with no quantitative comparisons among the schemes. It is suggested to add bar charts with quantitative metrics to more clearly illustrate and quantify performance differences across schemes.

Response: Thank you very much for this helpful suggestion. Following your recommendation, we added additional quantitative comparisons to complement the zonal-mean line plots. Instead of introducing additional bar charts, we summarized the overall performance differences among the schemes using a new quantitative table (Table 3), which reports the annual mean GPP bias relative to FLUXNET for each parameterization scheme and MODIS. We believe that this tabular format provides a more concise and direct comparison of the integrated biases across schemes while avoiding redundancy with the existing zonal plots.

Specifically, we added the following text to the Results section:

The bias statistics further support these zonal characteristics (Table 3). The no-ozone case shows an overall annual mean bias of 284 gC m^{-2} yr^{-1} relative to FLUXNET. Among the ozone-stress simulations, the Li parameterization scheme produces the smallest annual mean bias (420 gC m^{-2} yr^{-1}), followed by the Li framework combined with Lombardozzi and Sitch thresholds and functions, with biases of 425 and 428 gC m^{-2} yr^{-1}, respectively, whereas the Lombardozzi parameterization scheme exhibits the largest bias (566 gC m^{-2} yr^{-1}). In contrast, MODIS shows the smallest annual mean bias (229 gC m^{-2} yr^{-1}) relative to FLUXNET.

In addition, we introduced the following new table:

Table 3. Annual mean GPP bias relative to FLUXNET across different ozone parameterization schemes.

Experiment name	I2000 (no ozone stress)	Li	Li with Lombardozzi thresholds and function	Li with Sitch thresholds and function	Lombardozzi	MODIS
Bias (gC m^{-2} yr^{-1})	284	420	425	428	566	229

3. It is recommended to further incorporate differences among biomes and potential directions for improvement into the discussion to enhance its depth.

Response: Thank you very much for this valuable suggestion. Following your recommendation, we further expanded the discussion of biome-dependent differences and incorporated more detailed implications for future model improvement in Section 5 ("Implications for future improvement in earth system models").

The following text was added to the Implications for future improvement in Earth system models Section:

The significant differences in GPP bias across biomes reveal key directions for improving ozone-stress parameterization in models. In biomes with intense radiation, such as wetland grasslands and tropical moist broadleaf forests, models generally underestimate GPP, especially with the Lombardozzi scheme. Therefore, regional calibration of ozone flux thresholds and response functions is needed for these areas. In contrast, in low-stress regions such as tundra, temperate grasslands, and forests, the Li scheme, with optimized thresholds and decoupled response functions, performs best, suggesting that more refined model adjustments are required in these regions. For dry and desert biomes, a slight overestimation is observed, especially in the no-ozone simulations, indicating that the model's performance is generally good for these regions but still leaves room for further improvements when accounting for other environmental factors. In conclusion, optimizing the model to suit the characteristics of different biomes will improve the accuracy of ozone stress simulations, thereby enhancing the reliability of climate predictions.

4. In addition, the Li scheme does not consistently perform best for crop types and under autumn conditions. This should also be reflected in the results and discussion to ensure that the conclusions are more comprehensive and objective.

Response: Thank you very much for this important suggestion. Following your recommendation, we revised the manuscript to provide a more balanced and objective evaluation of the Li scheme across different vegetation types and seasons. In the revised discussion and implications section, we explicitly clarified that, although the Li scheme generally performs best across most vegetation types and evaluation metrics, it does not consistently outperform the other schemes for crop types and under autumn conditions. We also emphasized that the NSE values of all ozone-stress parameterization schemes remain relatively low in summer and that the fit to FLUXNET can deteriorate after ozone stress is introduced.

To reflect these limitations more clearly, we added the following text to Section 5 ("Implications for future improvement in earth system models"):

In addition, although the Li scheme performs well across most vegetation types, especially shrubs, where it significantly reduces bias (as shown in Fig. 7), it does not always perform best in crop types and under autumn conditions. Particularly in summer, the NSE values for all ozone stress schemes are relatively low, and after ozone stress is introduced, the model's fit to FLUXNET deteriorates. This indicates that, although the Li scheme performs well in certain biomes and seasons, further optimization is still needed for crop types and seasonal variations.

Reviewer: 2
Comments to the Author

I enjoyed reading this manuscript providing a deep dive into alternative ozone damage functions and their performance. The original difference between the Sitch and Lombardozzi approach principally related to whether there is a decoupling in the impact of ozone on photosynthesis and stomatal conductance, e.g. there was empirical evidence of a decoupling under water stress, i.e. sluggish stomata. In the Sitch model the global parameter “a” (eqn 4) was calibrated for that DGVM and monthly chemistry-modelled concentration fields to reproduce the empirical dose-response function for a high and low sensitivity species within a plant functional type grouping. I would anticipate implementation of Sitch et al. scheme into another DGVM would require a recalibration of this parameter (can you confirm whether or not this was done?). For example, what was really done in the runs: Li + Lombardozzi thresholds and functions and Li+Sitch thresholds and function? (any retuning?) I like the Li approach as it considers the effect of cumulative dose on phenology/leaf turnover. Given Li “adopts a data-driven optimal threshold Y” and best form of equation to fit the data, one would anticipate its best performance (especially as the combination approaches may not have been retuned)? Nevertheless, it is interesting to see the comparison of performance against multiple datasets (the comparison was thorough). Given the above I am particularly interested in the comparison of Li and Lombardozzi parameterization schemes. Overall there are some interesting elements to this study and warrants publication although the justification and explanation around inclusion of the mixed schemes should be elaborated.

Response: We sincerely thank you for your highly professional and insightful comments. We especially appreciate your in-depth discussion regarding the mechanistic differences among the Sitch, Lombardozzi, and Li ozone-stress parameterization schemes, particularly concerning the decoupling between photosynthesis and stomatal conductance, as well as the necessity of recalibration when transferring parameterization schemes across different DGVM/LSM frameworks. Following your suggestions, we revised the manuscript to clarify the scientific rationale and limitations of the mixed-parameterization experiments. Specifically, we now explicitly state in the Abstract, Introduction, Discussion, and Conclusion that the mixed experiments using thresholds and response functions derived from the Sitch and Lombardozzi schemes were conducted without additional recalibration or retuning within the Li framework. Therefore, these experiments should primarily be interpreted as structural-sensitivity experiments rather than as fully optimized implementations of those parameterizations within the Li framework. In addition, we further expanded the Discussion section to explain in greater detail the rationale for introducing the mixed schemes and to strengthen the comparison of the mechanistic differences between the Li and Lombardozzi parameterization schemes. We also incorporated experimental evidence from previous ozone exposure studies to demonstrate that ozone flux thresholds remain highly uncertain across vegetation types and species, thereby further justifying the need to conduct structural comparison experiments within a unified framework. In the revised manuscript, modifications made in response to your comments are highlighted in **red**, while revisions associated with another reviewer are highlighted in **blue** for clarity. We sincerely appreciate your valuable comments and guidance, which have substantially improved the quality, rigor, and readability of the manuscript.

Specific comments:

1. I would anticipate implementation of Sitch et al. scheme into another DGVM would require a recalibration of this parameter (can you confirm whether or not this was done?). For example, what was really done in the runs: Li + Lombardozzi thresholds and functions and Li+Sitch thresholds and function? (any retuning?)

Response: Thank you very much for this important comment. We agree that transferring the Sitch parameterization scheme into another DGVM/LSM framework would require recalibration of the empirical parameters, including the global parameter “a”, because the original parameterization was calibrated under a specific model structure and chemistry-forcing configuration.

However, in this study, we did not perform additional retuning or recalibration when implementing the mixed experiments (Li + Lombardozzi thresholds/functions and Li + Sitch thresholds/functions). Instead, we only replaced the ozone flux thresholds and response-function formulations within the Li framework while keeping all other model structures and parameter settings unchanged. Therefore, the primary purpose of these mixed experiments was to serve as structural sensitivity experiments to isolate the effects of threshold definitions and response-function forms, rather than the influences arising from other structural components of the parameterizations.

To clarify this point, we have now explicitly added corresponding statements in the Abstract, Introduction, Discussion, and Conclusion sections.

The following text has been added to the Abstract:

The mixed experiments using thresholds and response functions derived from the Sitch and Lombardozzi schemes were implemented without additional recalibration, allowing structural sensitivities to be evaluated consistently within the Li framework.

The following text has been added to the Introduction:

Notably, the mixed parameterization experiments using thresholds and response functions derived from the Sitch and Lombardozzi schemes were not recalibrated within the Li framework and should therefore be interpreted primarily as structural sensitivity experiments.

The following text has been added to the Conclusion:

Notably, the mixed experiments incorporating thresholds and response functions derived from the Sitch and Lombardozzi schemes were conducted without additional recalibration; therefore, the differences reported here primarily reflect structural sensitivities rather than fully optimized implementations of those parameterizations within the Li framework.

2. Overall there are some interesting elements to this study and warrants publication although the justification and explanation around inclusion of the mixed schemes should be elaborated.

Response: Thank you for this insightful suggestion. Following your comment, we substantially expanded the discussion better to justify the inclusion of the mixed parameterization experiments and to elaborate on the mechanistic differences between the Li and Lombardozzi schemes.

Specifically, we added a new discussion section comparing the ozone flux thresholds adopted in the Sitch, Lombardozzi, and Li schemes with evidence from previous ozone exposure experiments across different vegetation types. The revised manuscript now discusses how experimental studies suggest substantial uncertainty in ozone sensitivity among needleleaf forests, broadleaf forests, shrubs, grasses, and crops. For example, some experiments indicate stronger ozone sensitivity in loblolly pine than in yellow-poplar, while others show that crops such as soybean can be substantially more sensitive than woody broadleaf species. These observations are not always fully consistent with the threshold assumptions adopted in existing parameterization schemes.
We also clarified that one important motivation for introducing the mixed schemes was to evaluate the structural sensitivity associated with threshold selection and response-function forms under a unified Li framework. In addition, by comparing the Li + Lombardozzi experiment with the original Lombardozzi scheme, we were able to isolate further the role of cumulative memory effects from threshold and response-function differences alone.

The following text has been added to the Discussion:
Experimental evidence from previous ozone exposure studies also suggests that the ozone flux thresholds currently adopted across different parameterization schemes remain highly uncertain across vegetation types. For example, experiments comparing loblolly pine (needleleaf) and yellow-poplar (broadleaf) showed stronger ozone sensitivity in the needleleaf species (Tjoelker et al., 1991), whereas studies on Magnolia denudata and Cotinus coggygria indicated that broadleaf trees could be more ozone-sensitive than shrub species (Xu et al., 2021),. Other controlled experiments further suggested a sensitivity gradient of soybean > tobacco > poplar, implying that crop species are generally more sensitive to ozone than many broadleaf woody species (Dai et al., 2019). However, the Lombardozzi parameterization scheme applies a uniform threshold of 0.8 nmol m^{-2} s^{-1} to all five vegetation categories, which may oversimplify species- and PFT-dependent detoxification capacity. Similarly, the original Sitch scheme was developed under limited observational constraints, lacking dedicated experimental evidence for shrub and grass vegetation types; therefore, shrub thresholds were assumed to be equivalent to those for broadleaf vegetation, while grass thresholds were assumed to be similar to those for crops. In contrast, the Li scheme assigns higher ozone sensitivity to needleleaf forests than to broadleaf forests, a result that is not always consistent with the experimental evidence summarized above. Taken together, these inconsistencies indicate that ozone flux

thresholds remain insufficiently constrained and still require further recalibration across vegetation types and model frameworks.

This uncertainty was one of the primary motivations for introducing the mixed parameterization experiments in this study. By combining the Li framework with alternative threshold and response-function settings from the Lombardozzi and Sitch schemes, we aimed to isolate the effects of threshold definitions and response-function forms from other structural components of the parameterizations. Importantly, we did not perform additional retuning or recalibration of the empirical parameters, including the global parameter originally used in the Sitch scheme, when transferring these thresholds and response functions into the Li framework. Therefore, the mixed experiments should be interpreted as structural-sensitivity experiments rather than as optimized implementations. In addition, the comparison between the Li+Lombardozzi and the original Lombardozzi schemes enabled us to distinguish further the role of cumulative memory effects from that of threshold and response-function choices alone. Our results suggest that differences in long-term ozone accumulation and canopy memory can substantially alter the magnitude and spatial pattern of simulated ozone damage, even under similar threshold settings.