Dear Reviewer 1 and Topical Editor:

On behalf of my co-authors, we thank you very much for reviewing our manuscript and giving us highly constructive comments and suggestions. We deeply appreciate your feedback on our manuscript entitled "*Sensitivity and Uncertainty Analysis of China's Terrestrial Carbon-Water Cycle Using a Dynamic Global Vegetation Model*" (egusphere-2025-6076).

This document contains our detailed, point-by-point responses to all the comments raised by **Reviewer 1**. In these responses, we have addressed your concerns and clearly outlined the comprehensive modifications we plan to implement in the formal revised manuscript.

We firmly believe that your insights will greatly improve the biogeoscientific relevance and overall quality of our study. Looking forward to hearing from you.

Best regards,

Fulai Feng

School of Geography and Tourism, Shaanxi Normal University

No. 620, West Chang'an Street, Chang'an District, Xi'an 710119, China

Response to the Reviewers #1

RC1: Comment on egusphere-2025-6076

This discussion paper presents a global sensitivity analysis of the LPJ-GUESS model that also considers changes in sensitivities induced by different climatic zones across China. The overall framing, conceptualization and presentation of the manuscript are very good. The language is easily accessible; the questions and methods are clear and relevant for the vegetation modelling community, the figures are nice, and the general logic of the manuscript is clearly understandable. The topic fits well within the scope of Biogeosciences (BG), but given the technical nature of the manuscript, it would also fit within the scope of Geoscientific Model Development (GMD).

Response: We sincerely appreciate your positive assessment of our manuscript and your constructive feedback. Your deep understanding of dynamic global vegetation modeling and sensitivity analysis has provided us with invaluable insights. We have carefully considered all your major concerns, particularly regarding parameter ranges, reproducibility, and the interpretation of spatial sensitivity patterns. A point-by-point response to your comments is detailed below.

[Major Comment 1] The motivation for the parameter ranges should be better explained and the current ranges should possibly be changed. Parameter ranges should always be informed by ecological plausibility. If nothing is known about the effect of a parameter, an initial manual sensitivity analysis may be helpful to determine appropriate ranges. This is in any case better than setting parameters to technically possible min/max ranges or ad-hoc relative values such as +/- 25%. While important to resolve, I do not expect this to have major impacts on the results.

Response: We fully agree with the reviewer's fundamental perspective that parameter ranges in ecosystem modeling should ideally be constrained by ecological plausibility and empirical observations whenever such information is available. As the reviewer correctly pointed out, using technically permissible bounds or ad hoc relative deviations may introduce subjectivity and can potentially push the model into biologically unrealistic states.

At the same time, we would like to clarify that the use of standardized relative perturbations around default parameter values is a widely adopted exploratory practice in the initial global sensitivity analysis (GSA) of complex process-based models, particularly when comprehensive empirical prior information is unavailable for all parameters. For example, Li et al. (2022) explicitly stated that "the upper and lower limits of the parameters were set to 10% of the model's default value" and further noted that this choice was consistent with previous sensitivity-analysis settings adopted in models such as DSSAT-CERES-Wheat and CROPGRO (e.g., Song et al., 2014; Li et al., 2020; Wu et al., 2015). In the same study, Li et al. (2022) also pointed out that Zhang et al. (2011) allowed some physiological and ecological parameters in the BIOME model to vary by 20–50% around their default values during parameter optimization. Likewise, preliminary sensitivity assessments based on fixed relative perturbations have long been used in model analysis studies. For instance, Miller (1974) evaluated the sensitivity of canopy conductance and decoupling to a ±10% parameter change, and Bartelink (1998) quantified relative sensitivity by increasing and decreasing parameter values by 5% around their defaults.

Therefore, our use of a ±25% range around default values should be understood as an exploratory structural-diagnostic choice that is broadly consistent with relative perturbation strategies reported in the literature. Our original intention was not to claim that these ranges are strictly ecologically realistic for all parameters, but rather to diagnose the model's structural sensitivity across a broad yet controlled mathematical space prior to any formal data-constrained calibration.

Nevertheless, we fully accept the reviewer's concern that such exploratory ranges should not be conflated with ecologically justified parameter bounds. To address this point more explicitly in the revised manuscript, we will revise both the Methods and Discussion sections. First, we will clarify in the Methods that the adopted parameter ranges represent an exploratory screening design informed by precedent in the sensitivity-analysis literature, rather than a definitive ecological parameterization scheme. Second, we will add a dedicated paragraph in the Discussion (Section 4.3) to acknowledge that the sensitivity patterns identified here are conditional on the assumed parameter bounds, and that future regional calibration should rely on observationally constrained trait databases and parameter-specific empirical evidence to define ecologically plausible ranges. Accordingly, the highly sensitive parameters identified here should be regarded primarily as priority candidates for future empirical constraint and regional calibration, rather than as direct justification for immediate parameter adjustment.

References:

Li, Y., Wang, Y., Sun, Y., and Li, J.: Global sensitivity analysis of the LPJ model for Larix olgensis Henry forests NPP in Jilin Province, China, Forests, 13, 874, https://doi.org/10.3390/f13060874, 2022.

Zhang, T., Sun, R., Hu, B., and Feng, L.: Using simulated annealing algorithm to optimize the parameters of Biome-BGC model, Chin. J. Ecol., 30, 408–414, 2011.

Bartelink, H. H.: A model of dry matter partitioning in trees, Tree Physiol., 18, 91–101, https://doi.org/10.1093/treephys/18.2.91, 1998.

Miller, D. R.: Sensitivity analysis and validation of simulation models, J. Theor. Biol., 48, 345–360, 1974.

Li, B., Li, C., Yao, M., Wei, X., Bao, Z., and Sun, X.: Sensitivity and uncertainty analysis for CROPGRO-Tomato model at different irrigation levels, J. Shenyang Agric. Univ., 51, 153–161, 2020.

Wu, L., Zhang, F., Fan, J., Zhou, H., Xing, Y., and Qiang, S.: Sensitivity and uncertainty analysis for CROPGRO-cotton model at different irrigation levels, Trans. Chin. Soc. Agric. Eng., 31, 55–64, 2015.

Song, M., Feng, H., Li, Z., and Gao, J.: Global sensitivity analyses of DSSAT-CERES-Wheat model using Morris and EFAST methods, Trans. Chin. Soc. Agric. Mach., 45, 124–166, 2014.

[Major Comment 2] In the data availability statement, the authors provide download links for model code and data, but this would likely not be sufficient to reproduce the results, as neither the analysis code, nor the modified model code are provided (I assume the LPJ-GUESS code must have been modified to change the hard-coded parameter values the authors refer to). For a revision, I suggest the authors should link to a repository that contains their entire project, including (as far as copyrights allow it) their modified model code, the data and their analysis code (e.g. the code to calculate sensitivity indices as well as the code used to generate the figures), as well as the intermediate results (i.e. the sensitivity indices) that were used to generate the figures. Please make sure that the results are computationally reproducible, e.g. by appropriately fixing random seeds before running the sensitivity analysis.

Response: We deeply appreciate your rigorous standard for computational reproducibility, which we agree is essential for the modeling community. We have fully adopted your suggestion. Instead of merely providing download links in the revised manuscript, **we have already created a comprehensive, publicly accessible repository on Zenodo (DOI: https://doi.org/10.5281/zenodo.19533307)**.

To ensure complete transparency and address all of your specific requests, the repository is strictly organized into the following logical workflow:

1. Modified Model Source Code: Includes the customized LPJ-GUESS C++ source code (in full compliance with the original Mozilla Public License v2.0) used to expose the hard-coded physiological parameters.

2. Analysis Scripts: Contains all Python scripts logically numbered by workflow step. This includes parameter sampling, automated parallel model execution, sensitivity index calculation (Morris, eFAST, Sobol'), and figure generation.

3. Input Data & Configurations: Provides the exact parameter bounds and site configurations used in our study.

4. Intermediate Results: We have provided the fully calculated sensitivity indices. This allows readers and reviewers to directly execute our plotting scripts and reproduce the figures in the manuscript instantly, without needing the massive computational resources required to re-run the entire DGVM.

We have comprehensively updated the *Data Availability Statement* in the revised manuscript to reflect this new, fully reproducible repository. We believe this transparency will not only satisfy reproducibility requirements but also allow other LPJ-GUESS users to easily adapt our GSA framework for their own research.

[Major Comment 3] An interesting but also a bit unusual aspect of this study is that the authors used and compared several sensitivity indices in parallel. This is interesting, both to learn about the differences between the indices, and to explore the robustness of the conclusions, but I would have liked to see more analysis and discussion about the reasons for differences where they occur. Also, I had concerns about the standardization that is employed and would ask the authors to consider removing it (see detailed comments).

Response: We sincerely appreciate this highly insightful statistical comment, which also ties directly into your detailed comments regarding Line 88 and Line 274. We completely agree with your assessment regarding the fundamental differences between these mathematical paradigms and the inherent flaws in our previous cross-method standardization approach. We have comprehensively revised our methodology and visual presentation across the entire manuscript to address these concerns.

1. Rationale for Parallel Multi-Method Application (Addressing L88):

The primary motivation for employing three methods in parallel was to establish methodological robustness and cross-validation. In the DGVM community, the immense computational cost of running process-based models often prohibits the use of rigorous variance-based methods (like Sobol') for large parameter sets. By executing them in parallel, we aimed to demonstrate that a computationally cheaper screening method (Morris) can reliably identify the same top-tier dominant parameters (e.g., *ALPHA_C3*, *common_reprfrac*) as the highly expensive Sobol' method, even if minor rank shuffling occurs among less sensitive parameters. In the revised manuscript, we will make this explicit in the Introduction and Methods sections.

2. Discussion on the Reasons for Differences:

We agree that the manuscript lacked a deep discussion on why the indices sometimes diverge. As you astutely pointed out, these methods rely on fundamentally different mathematical paradigms. The Morris method is essentially derivative-based (calculating the mean of elementary effects, μ, along trajectories), which captures the average local sensitivity across the parameter space. In contrast, eFAST and Sobol' are variance-based, decomposing the total output variance into fractional contributions. Therefore, a parameter that causes sharp, localized, non-linear jumps in model output might register a high elementary effect in Morris, but if these jumps only occur in a rare subset of the parameter space, its overall contribution to the total output variance (Sobol' S_{T_i}) might be diluted. We will add a specific paragraph in Sections 3.3 and 4.1 to explicitly discuss these mathematical reasons, clarifying that these divergences reflect different facets of model behavior rather than errors.

3. Removal of Standardization and Systematic Visual Overhaul (Addressing L274):

Your criticism regarding the "Combined Score" standardization is entirely valid. Mathematically, normalizing a variance-based proportion (Sobol'/eFAST, strictly [0,1]) and a derivative-based mean effect (Morris μ, which is scale-dependent) to a shared scale and averaging them conceptually conflates different statistical properties.

We have completely removed this numerical standardization procedure for parameter selection and updated all related visualizations (Figures 2, 4, and 5) to reflect the raw statistical truths:

Implementation of Non-parametric Ranking (Figure 2): Instead of using a normalized composite score, we now rely exclusively on non-parametric rank-based metrics. Since a parameter's ordinal rank is invariant to the absolute scale or unit of its underlying index, calculating the Average Rank across the three methods is a strictly rigorous way to identify the top consensus parameters. The x-axis ordering in the revised Figure 2 is now strictly determined by this Average Rank.

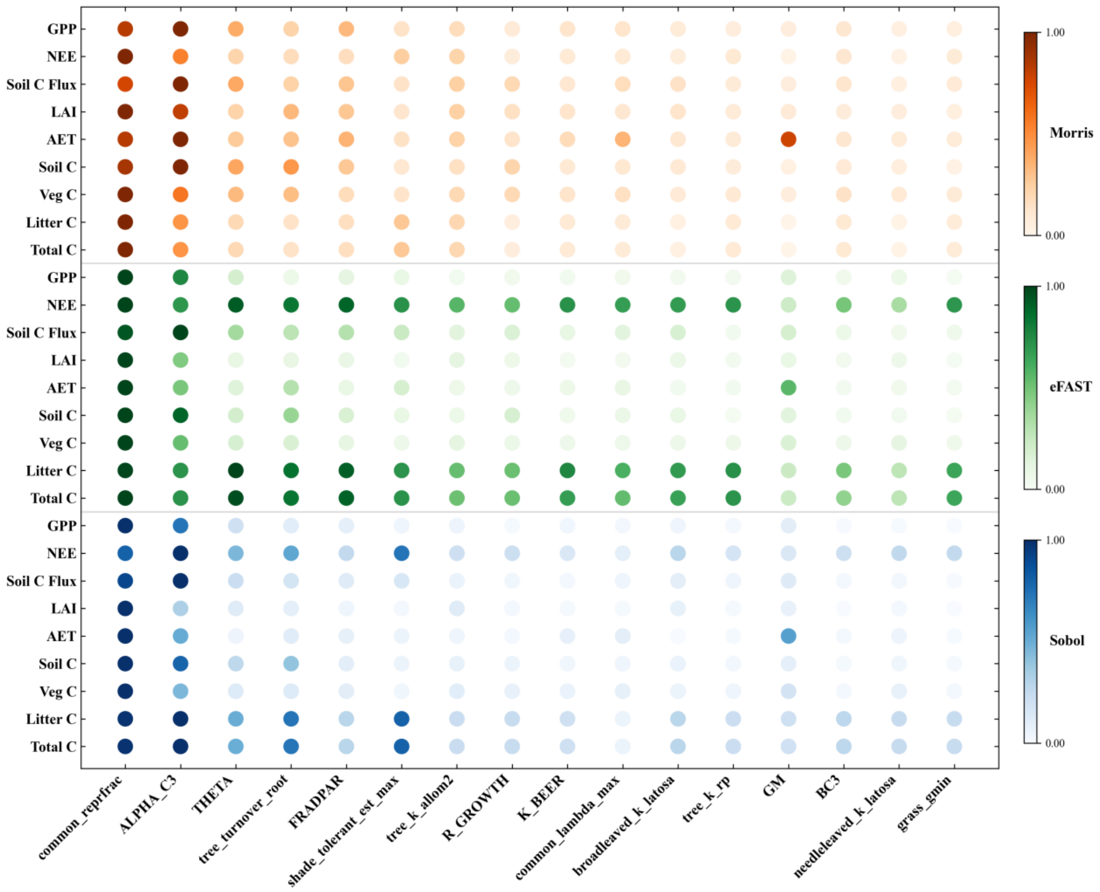


Fig. 2: Heatmap of the top 10 most sensitive parameters across all target variables identified by the Morris, eFAST, and Sobol' methods. ***Note on visualization:*** **To maintain visual readability across target variables with vastly different absolute magnitudes (particularly for the scale-dependent Morris μ index), the color intensities in this heatmap are row-standardized (scaled to a maximum of 1.0 within each row). However, the selection of parameters and their ordering along the x-axis remain strictly determined by non-parametric Average Ranks derived from the raw, unstandardized indices, ensuring that the visual scaling does not affect the statistical conclusions.**

Updating Consistency Analysis (Figure 4): In the revised Figure 4, the diagonal bar charts now display the true raw sensitivity indices rather than normalized scores, while the off-diagonal scatter plots continue to rigorously demonstrate cross-method consistency using the non-parametric Spearman's rank correlation (ρ).

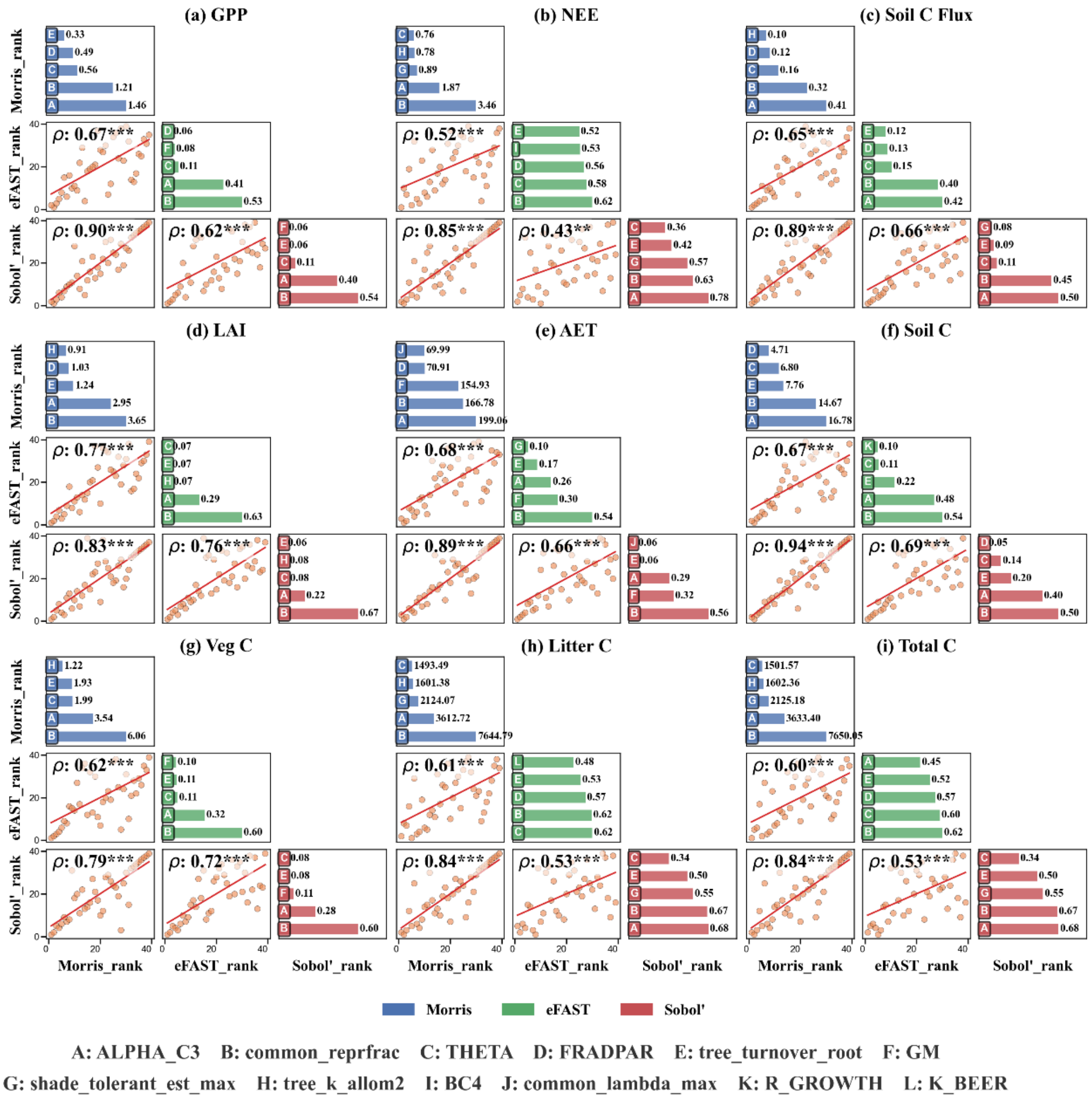


Fig. 4: Consistency analysis of parameter sensitivity rankings identified by Morris, eFAST, and Sobol' methods across nine output variables. Each of the nine large panels corresponds to a specific model output variable. Within each large panel, the sub-panels compare the sensitivity rankings of the three methods: the diagonal sub-panels display the raw (unstandardized) sensitivity indices for the top-five ranked parameters from each method; the lower-left sub-panels show scatter plots comparing the parameter sensitivity ranks between pairs of methods, annotated with the Spearman's rank correlation coefficient ρ and its statistical significance (* represents p < 0.001, ** represents p < 0.01). The solid line represents the linear regression fit to the ranked scatter points.**

Restoring Physical Meaning in Line Plots (Figure 5): In the revised Figure 5, we have restored the strict physical meaning of the variance-based methods. The y-axes for the eFAST and Sobol' panels now directly display the raw Total Sensitivity Index (S_{T_i}), correctly showing the absolute fraction of variance explained (which naturally caps below 1.0, unlike our previous artificially stretched plots). For the Morris method, the derivative-based μ values were scaled to their maximum within each output variable (Relative μ) strictly to allow visual comparison of parameter ranking trends.

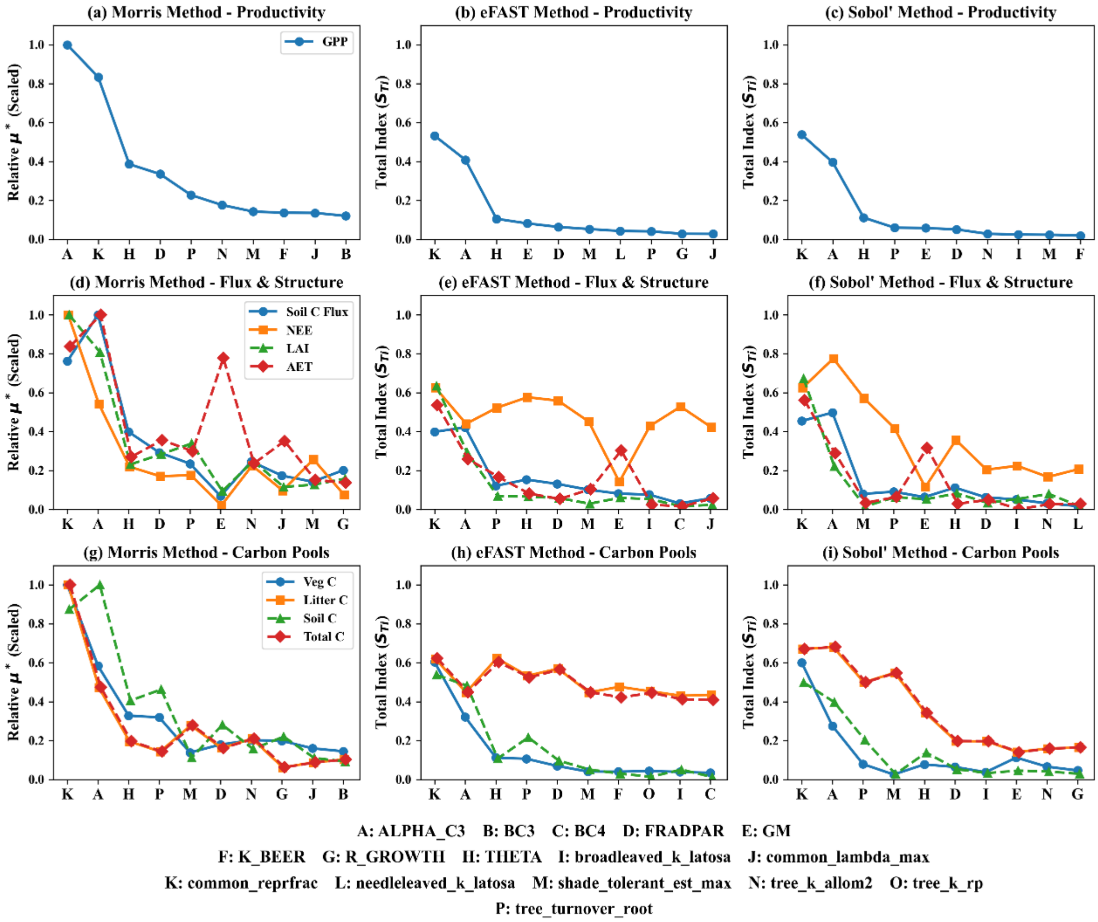


Fig. 5: Comparison of sensitivity parameter dynamics among the Morris, eFAST, and Sobol' methods. Each subplot shows the sensitivity index (Y-axis) of different model parameters (X-axis, letter codes) for three groups of target variables. The Y-axis displays the raw Total Sensitivity Index (S_{T_i}) for eFAST and Sobol', and the relative scaled mean effect (μ) for the Morris method. The rows represent the variable groups of Productivity, Flux & Structure, and Carbon Pools, respectively; the columns correspond to the three analysis methods.

4. Visual Proof of the Figure 2 Plotting Strategy (For Reviewer's Reference):

Following your advice, we reconsidered the visualization of the sensitivity indices in Figure 2. We found that plotting the strictly raw, unstandardized Morris μ values alongside variance-based indices (S_{T_i}) creates a significant visual compression issue.

Because μ is scale-dependent, its absolute values for massive carbon pools (e.g., Total C) are several orders of magnitude larger than those for annual fluxes (e.g., GPP), rendering the sensitivities of flux variables visually imperceptible on a single global color scale.

To resolve this while maintaining strict statistical rigor, we adopted a row-wise visual scaling approach for the revised Figure 2. Specifically, the color intensities are scaled to a maximum of 1.0 within each specific target variable row. This ensures that the relative importance of parameters is visually clear for every output variable, without misleading numerical conflation across different physical units.

To ensure absolute transparency, we have added the following clarification to the Figure 2 caption:

"Note on visualization: To maintain visual readability across target variables with vastly different absolute magnitudes (particularly for the Morris μ *index), the color intensities in this heatmap are row-standardized. However, the selection of parameters and their ordering along the x-axis remain strictly determined by non-parametric Average Ranks derived from the raw, unstandardized indices, ensuring that the visual scaling does not affect the statistical conclusions."*

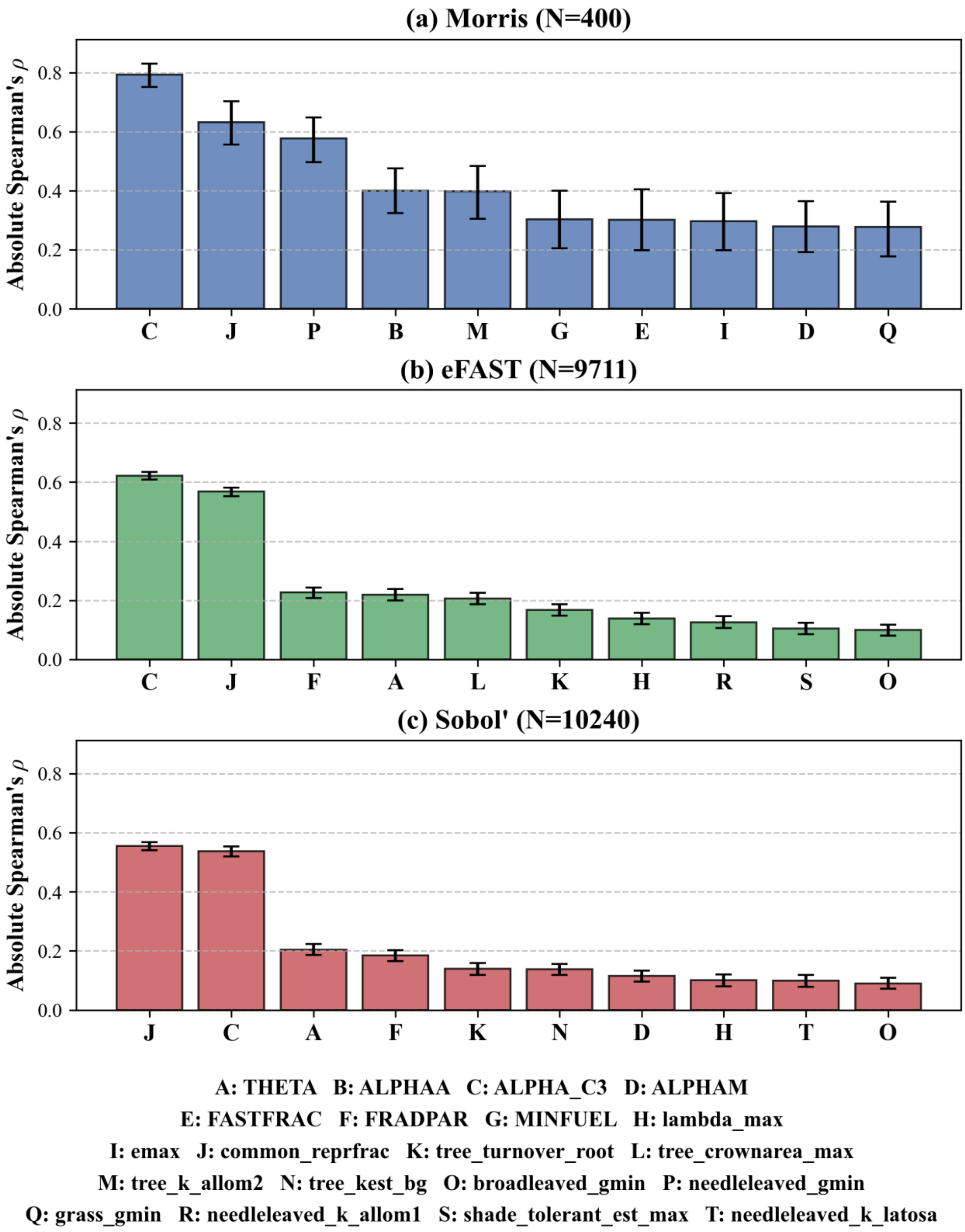
[Major Comment 4] The number of parameter draws / model runs to calculate the sensitivity indices seems to be worryingly low. In particular Sobol' indices often require a much larger number of iterations to converge, but the sample size for the Morris screening also seems rather low. It is critical that the authors demonstrate that the results are stable. This could be done either by repeating the analysis several times and showing that results are qualitatively identical, or by bootstrapping the existing simulations (which would be less computationally demanding).

Response: We highly appreciate the reviewer's rigorous attention to the statistical convergence of our sensitivity indices. We acknowledge that for a highly complex model with 39 parameters (D=39), our base sample sizes (Morris: N=400; eFAST: N=9711; Sobol': N=10240) might theoretically appear borderline to some readers for capturing high-order interactions.

However, as you perceptively noted, the computational cost of running a process-based DGVM (e.g., LPJ-GUESS with 500-year spin-ups at 13 sites) is immense, making a massive replication of simulations computationally prohibitive. Therefore, we enthusiastically adopted your excellent suggestion to perform a Bootstrap-based stability analysis on our existing extensive simulation outputs to rigorously prove convergence for all three methods.

Specifically, we bypassed the black-box constraints of traditional sensitivity software by developing a custom Python script to perform 1000 independent bootstrap resamples (with replacement) on the raw input-output data pairs for each of the three sampling designs (Morris, eFAST, and Sobol'). For each resample, we calculated the absolute Spearman's rank correlation (ρ) to quantify parameter dominance.

As explicitly demonstrated in the newly added Figure S1 (which now includes panels for all three methods), the 95% Confidence Intervals (represented by the error bars) for the top-ranking sensitive parameters are exceptionally narrow across the board. Regardless of the resampling permutations, the core parameters (e.g., *ALPHA_C3* and *common_reprfrac*) consistently maintain their dominant rankings with negligible variance in all three methods. This provides ironclad statistical evidence that our sample sizes have reached a robust state of convergence, and our core conclusions regarding parameter dominance are completely stable. We have integrated this validation step into Section 2.6.3 and added the corresponding convergence plot to the Supplementary Information.


(a) Morris (N=400)

(b) eFAST (N=9711)

(c) Sobol' (N=10240)

(Caption for Figure S1: Bootstrap stability check for parameter sensitivity across the three applied methods: (a) Morris, (b) eFAST, and (c) Sobol'. The bars represent the mean absolute Spearman's ρ derived from 1,000 bootstrap resamples of the original simulation outputs (N=400, N=9711, and N=10240, respectively). The remarkably narrow error bars (95% Confidence Intervals) across all panels indicate that the sensitivity rankings are highly stable and the utilized sample sizes are fully sufficient for robust convergence.)

[Major Comment 5] The authors conclude from the fact that there are regional changes in sensitivity that we need region-specific parametrizations of the model. This appears to be a misinterpretation of the results. If we would introduce regionalized model parameters, sensitivities would change between regions, but the fact that the authors find regional changes in sensitivity says nothing about the sense or regional model parameters. Rather, it says that the climatic drivers interact with the parameters in producing the outputs. This is also the interpretation in (Oberpriller et al., 2022), which is cited in the study.

Response: We highly appreciate this critical theoretical insight. You are absolutely correct, and we realize that our original phrasing inadvertently conflated "spatial variations in parameter sensitivity" with "the necessity for spatial variations in parameter values (regionalization)".

As you and Oberpriller et al. (2022) correctly point out, the spatial divergence in sensitivity we observed is fundamentally driven by the interactions between climatic boundaries (e.g., water availability, energy inputs) and the model parameters. In a water-abundant environment, the model's equations naturally render parameters related to water use efficiency less influential on the output variance. This demonstrates how environmental context dictates which parameters dominate the model uncertainty, but it does not inherently prove that the true biological trait values must differ across regions.

We will thoroughly correct this misinterpretation throughout the revised manuscript. We will shift our narrative from advocating for "region-specific parametrizations" to advocating for "context-dependent calibration priorities". For example, in the Abstract and Section 4.3 (Outlook), we will revise statements like: Original: "*...demand region-specific parameter sets that reflect their unique limiting factors.*" Revised: "*...demand context-aware calibration strategies. Because climatic drivers strongly interact with parameter sensitivities, efforts to reduce model uncertainty must prioritize constraining different sets of parameters depending on the regional environment (e.g., prioritizing allocation parameters in arid zones versus photosynthetic capacity in humid zones), as similarly noted by Oberpriller et al. (2022).*"

References:

Oberpriller, J., Herschlein, C., Anthoni, P., Arneth, A., Krause, A., Rammig, A., Lindeskog, M., Olin, S., and Hartig, F.: Climate and parameter sensitivity and induced uncertainties in carbon stock projections for European forests (using LPJ-GUESS 4.0), Geosci. Model Dev., 15, 6495–6519, https://doi.org/10.5194/gmd-15-6495-2022, 2022.

[Major Comment 6] Overall, the discussion of the results seemed in part very speculative to me (see detailed comments below). I would ask the authors to consider if their conclusions in the discussion are directly supported by the results generated in the sensitivity analysis.

Response: We accept this constructive criticism. We recognize that bridging the gap between a pure mathematical sensitivity analysis and real-world ecological mechanisms requires caution. Because our study is an uncalibrated, model-only diagnostic experiment, our findings primarily reflect the internal structural logic and equation behavior of the LPJ-GUESS model, rather than independently verified empirical ecological truths.

In the revised manuscript, we will carefully review Section 4 (Discussion) to "tone down" speculative statements and ensure our interpretations remain strictly within the bounds of what the GSA data supports. We will consistently qualify our ecological interpretations by framing them within the model's context. For instance, instead of stating "ecosystems in arid regions are controlled by resource allocation," we will revise it to state "within the model framework, simulated fluxes in arid regions are mathematically dominated by parameters governing resource allocation." By making these adjustments, alongside the new limitations paragraph mentioned in our response to Major Comment 1, we believe the revised discussion will be much more objective, rigorous, and directly anchored to our analytical results.

[Comment on L18 & L80] I’m not sure how people typically interact with the model. Don’t people also change parameters that are hard-coded in the source, or change the parameters included in the parameter files?

Response: Thank you for pointing out this ambiguity. While advanced developers certainly can and do modify the C++ source code, the vast majority of standard model users (especially in applied ecological studies) interact with LPJ-GUESS primarily through the external instruction (.ins) text files. These .ins files expose the PFT-specific traits, making them easily adjustable. Conversely, modifying the core physiological parameters requires altering the C++ source files and recompiling the entire model, which poses a technical barrier for many end-users. We will clarify this practical, operational distinction in the Introduction to better explain why "hard-coded" parameters are often overlooked in standard calibration efforts.

[Comment on L68] It’s not clear to me if you cite the reference as support for the statement, or as an example of a study which did pay attention to this. If the latter, cite this paper as: (but see XXX)

Response: We sincerely apologize for the ambiguity in our citation format. To clarify, the citation of Oberpriller et al. (2022) was intended to support our statement rather than serve as an exception. While their study is a highly comprehensive sensitivity analysis of the LPJ-GUESS model, they explicitly stated in their methodology that they exclusively "concentrated on carbon outputs (gross primary production... total standing biomass... and net biome productivity)" (Oberpriller et al., 2022). Thus, their work serves as a recent, prominent example of the prevailing trend: heavily focusing on the carbon cycle while leaving the integrated evaluation of coupled carbon-water outputs largely unaddressed.

To prevent any misunderstanding for the readers, we will rephrase this sentence in the revised manuscript to explicitly indicate it as a supporting example: "...with relatively limited attention to the integrated evaluation of carbon-water coupling processes (e.g., as seen in recent comprehensive analyses like Oberpriller et al., 2022)."

[Comment on L74] OK, the previous sensitivity analyses were conducted in Europe, but why would we expect that the results are different in China? Maybe you could make this more explicit by contrasting the climatic ranges of your study to previous studies (are you effectively considering a wider range of ecosystems?)

Response: This is an excellent suggestion. We do indeed expect different results because China encompasses a much steeper and wider hydro-thermal gradient than the relatively temperate and boreal focus of most European studies. China spans from tropical/subtropical monsoon climates in the southeast to extreme arid continental climates in the northwest, and includes the unique high-altitude alpine ecosystems of the Tibetan Plateau. We will make this explicit in the Introduction, emphasizing that this wider range of climatic extremes allows us to test the model's parameter sensitivities across a broader spectrum of environmental constraints.

[Comment on L81] Is this a new paragraph here? Seems sensible.

Response: Yes, this should be a new paragraph. We have corrected this formatting error.

[Comment on L88 & L240] I agree with the statement but what I don't understand is your motivation to run these 3 methods in parallel... If there is no reason to expect differences, it seems computationally wasteful...

Response: As detailed in our response to [Major Comment 3], our motivation for parallel execution was cross-validation. Given the high computational cost of DGVMs, we aimed to demonstrate whether a computationally cheaper screening method (Morris) could reliably identify the same dominant parameters as the highly expensive, variance-based Sobol' method in a high-dimensional parameter space. We will explicitly state this methodological cross-validation objective in the revised Introduction and Methods sections.

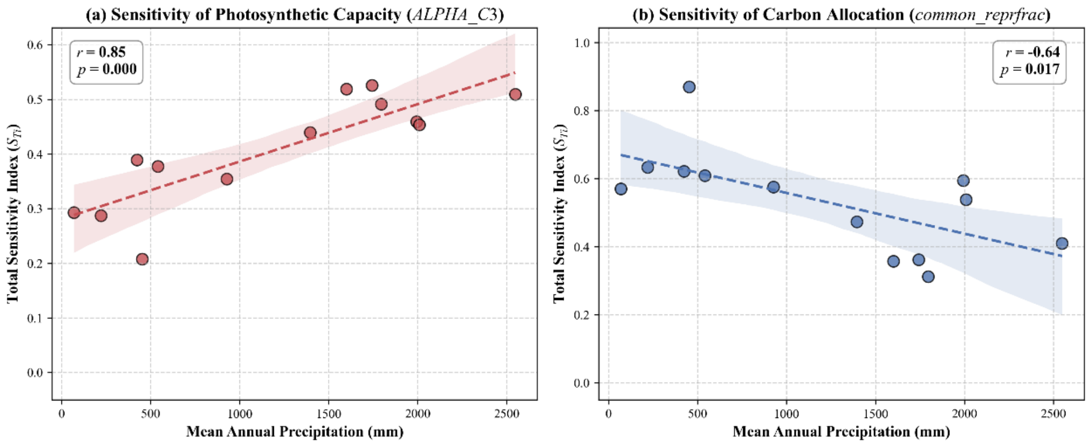
[Comment on L89 & L116] 13 Sites doesn't seem a lot. If they differ in multiple climate variables, I'm not sure how you will later disentangle which aspect of the climate is responsible for the change in sensitivity... What kind of representativeness are we aiming for here?

Response: We acknowledge that 13 sites are numerically few. However, these sites were not chosen randomly. As already detailed in Text S1 of our original Supplement ("Detailed Methodology for Study Site Selection"), these sites were rigorously identified through a two-stage spatial equilibrium sampling method using 23-year MODIS land cover dynamics and national nature reserve boundaries. They serve as highly representative "climatic and ecological archetypes."

Furthermore, to better disentangle the climatic drivers (as you rightly suggested), we have now explicitly quantified the climate space each site occupies using Mean Annual

Precipitation (MAP). As demonstrated in our newly added mechanistic analysis (**which is detailed in our response to Reviewer 2, and reproduced below for your convenience**), utilizing MAP allows us to mathematically link shifts in parameter sensitivity directly to this specific hydro-climatic gradient (i.e., transitioning from light limitation to water limitation), rather than just relying on geographic locations.

(Note: This new analysis will be formally added as Figure 9 in the revised manuscript).


(a) Sensitivity of Photosynthetic Capacity (*ALPHA_C3*)

(b) Sensitivity of Carbon Allocation (*common_reprfrac*)

Fig 9. Mechanistic shift in parameter dominance along the precipitation gradient. Scatter plots showing the Total Sensitivity Index (S_{T_i}) of GPP to (a) photosynthetic capacity (*ALPHA_C3*) and (b) carbon allocation (*common_reprfrac*) across the 13 archetypal sites, plotted against Mean Annual Precipitation (MAP). The shaded areas represent the 95% confidence intervals of the linear regression fits. The strong positive ($r = 0.85$, $p < 0.001$) and negative ($r = -0.64$, $p = 0.017$) correlations statistically demonstrate the model's transition from a photosynthesis-limited regime in resource-rich (humid) ecosystems to an allocation-limited regime in resource-limited (arid) ecosystems.

[Comment on L105] Why does it matter that the sites are undisturbed? Isn't the model only driven by climate? Would sensitivity calculations be affected by that?

Response: LPJ-GUESS is driven by climate, but it simulates potential natural vegetation dynamics unless explicit anthropogenic land-use forcing (e.g., crop harvest, pasture management) is provided. By selecting undisturbed nature reserves, we ensure that the actual observed vegetation (used to verify our PFT setups) theoretically aligns with the model's unmanaged simulations. If we selected heavily managed sites, the mismatch between modeled natural dynamics and actual disturbed states could introduce structural confounding factors into our interpretation of parameter sensitivity. We will briefly clarify this rationale in Section 2.1.

[Comment on L130] Driving or Driver?

Response: Corrected. We will change "Driver data" to "Driving data" or "Meteorological forcing" throughout the text.

[Comment on L183] I don't agree with this argument... parameters should be set within the range of plausible values...

Response: As extensively discussed in our response to [Major Comment 1], we fully concede this point. We will add a dedicated limitation section explicitly acknowledging that using technical boundaries (e.g., [0, 1] or ±25%) for exploratory diagnostics limits the immediate ecological applicability of the results, and that future regional calibrations must use trait databases to define biologically plausible ranges.

[Comment on L190] Uniform makes sense here, but similar to above, a uniform distribution does not reflect an absence of prior information, it makes the assumption that all values are equally likely a priori.

Response: We appreciate this precise statistical correction. We will revise the sentence to: "parameter values were drawn from a uniform distribution... assuming that all values within the specified bounds are equally likely a priori."

[Comment on L248] Adjectives such as "systematically", "comprehensive" here and elsewhere could be erased. Just say what you did and leave it to the reader to decide if they find it comprehensive.

Response: We completely agree. We will conduct a thorough review of the manuscript to remove these subjective adjectives and maintain a strictly objective tone.

[Comment on L263] How do you know it's sufficient? You could e.g. explore if results are stable when repeating the analysis...

Response: As detailed in our comprehensive response to [Major Comment 4], we have addressed this critical concern by implementing a rigorous Bootstrap resampling analysis (1,000 iterations for each of the three methods, bypassing internal software constraints). The extremely narrow 95% confidence intervals, now visualized in the new Figure S1 in the Supplement, provide concrete, statistical proof of the stability and absolute convergence of the parameter rankings, ensuring that our sample size is fully sufficient.

[Comment on L274] Not sure about the standardization of the sensitivity indices... Please consider if it could be removed without harming your analysis goals.

Response: As thoroughly addressed in our response to [Major Comment 3], we fully agree. We will completely remove the standardization procedure in the revised manuscript, present the raw indices on their native scales, and rely exclusively on non-parametric rank correlation metrics for cross-method comparisons.

[Comment on Fig. 4] ...What is also not clear to me is what sensitivities are you comparing here. eFast, for example, can produce total and direct / first order. Which of those is used for the comparison?

Response: We apologize for the omission. For both eFAST and Sobol', we utilized the Total Sensitivity Index (S_{T_i}) for the consistency comparisons with the Morris μ index, as S_{T_i} encompasses both main and interaction effects. We will explicitly state this in the figure caption and methodology section. Additionally, as part of our visual and methodological overhaul (detailed in Major Comment 3), Figure 4 has been completely updated. It now strictly uses unstandardized raw indices for the marginal bar charts and pure, rank-based Spearman's ρ for the scatter plots, fully resolving any standardization concerns.

[Comment on L570] For this entire section: I didn’t understand why this observation deserves an entire subsection... The reason for this pattern is that LPJ-GUESS, like virtually all process-based ecosystem models, has a largely multiplicative structure... To me, the result that absolute uncertainties are higher in areas that have large values seems expected.

Response: We concede that mathematically, this trade-off is an expected outcome of a multiplicative model structure. Our intention was not to present this as a novel mathematical discovery, but rather to explicitly map its spatial realization across China and highlight its practical implications for carbon budgeting (i.e., where observation networks should prioritize constraining fluxes vs. reducing relative model instability). In the revised manuscript, we will tone down the "surprising" nature of this finding. We will reframe this subsection to focus on the management and calibration implications of this expected mathematical behavior across diverse regional contexts.

[Comment on L609, L612, L663] See my general comment – this seems to be a big misunderstanding... This is no indication that you need a regional parameterization. These statements need to be changed.

Response: As detailed in our response to [Major Comment 5], we fully accept this theoretical correction. We have completely rewritten this narrative throughout the abstract, discussion, and conclusion. We no longer advocate for "region-specific parametrizations," but rather emphasize the need for "context-dependent calibration priorities," acknowledging that climatic drivers interact with parameters to dictate model behavior.

[Comment on L625] You could note that this is done in other sensitivity studies on LPJ-GUESS and briefly summarize the results of such an analysis.

Response: Thank you for the suggestion. We will briefly add a comparison in the Discussion (Section 4) summarizing how our parameter ranking results align with or diverge from previous LPJ-GUESS GSA studies (e.g., Pappas et al., 2013; Oberpriller et al., 2022).

References:

Oberpriller, J., Herschlein, C., Anthoni, P., Arneth, A., Krause, A., Rammig, A., Lindeskog, M., Olin, S., and Hartig, F.: Climate and parameter sensitivity and induced uncertainties in carbon stock projections for European forests (using LPJ-GUESS 4.0), Geosci. Model Dev., 15, 6495–6519, https://doi.org/10.5194/gmd-15-6495-2022, 2022.

Pappas, C., Fatichi, S., Leuzinger, S., Wolf, A., and Burlando, P.: Sensitivity analysis of a process-based ecosystem model: Pinpointing parameterization and structural issues, JGR Biogeosciences, 118, 505–528, https://doi.org/10.1002/jgrg.20035, 2013.

We appreciate your warm work earnestly, and hope that the correction will meet with approval. We tried our best to improve the manuscript and made some changes in the manuscript. These changes will not influence the content and framework of the paper.

Once again, thank you very much for your comments and suggestions.