Response to Anonymous Referee #1

This article is based on the multi-parameterization scheme of the Noah-MP land surface model and systematically evaluates the sensitivity and uncertainty of different physical processes on the simulation of gross primary productivity (GPP) in China's terrestrial ecosystems, which has important scientific significance and practical value. However, I suggest that the authors need to make a major revision to this study and that the study be accepted only after a thorough analysis of the sources of error and uncertainty.

We sincerely thank the reviewer for the positive assessment of the scientific significance and practical value of this study. We also fully agree that a more thorough analysis of the sources of model error and uncertainty is essential before drawing conclusions from the GPP sensitivity analysis.

In response, we have substantially revised the manuscript and reorganized the analysis to strengthen the diagnosis of model uncertainties. First, we clarified the research gaps and the specific contribution of this study in the Introduction, particularly regarding GPP uncertainty attribution across major terrestrial ecosystems in China. Second, we evaluated the potential uncertainty associated with land-cover representation by conducting additional simulations using two alternative vegetation/land-cover datasets, while keeping the forcing data and model configurations unchanged.

Third, we added spin-up convergence diagnostics for water, energy, carbon, and representative dynamic vegetation variables, including LAI, leaf mass, fine-root mass, stem mass, NPP, and total photosynthesis. These analyses demonstrate that the rapidly responding vegetation states and carbon-assimilation processes approached quasi-steady conditions during the final spin-up cycles.

Fourth, to identify the sources of poor site-level GPP performance, we added process-based diagnostic analyses using available flux-tower observations of GPP, APAR, soil moisture, latent heat flux, and sensible heat flux. These analyses show that low model performance at several grassland sites and the DHF forest site is associated with mismatches in canopy radiation absorption, soil-water dynamics, and surface-energy partitioning, which may propagate into GPP through water-stress and stomatal-regulation processes.

Finally, we revised the Discussion section to distinguish the uncertainty associated with forcing data, land-cover representation, spin-up, flux partitioning, PML GPP, and Noah-MP structural parameterizations. We also added ecosystem-specific and seasonal sensitivity analyses to better interpret the roles of radiation transfer, soil-moisture limitation, turbulence, and runoff parameterizations in GPP simulation uncertainty.

We believe that these revisions substantially improve the mechanistic interpretation, transparency, and robustness of the manuscript. Detailed responses to each comment are provided below.

Specific：

Q1. Although the study proposes the goal of 'identifying key physical processes,' the introduction does not fully explain the specific gaps in existing research on 'GPP sensitivity analysis in China.' The differences and innovations of this study compared to previous work would stand out more if the authors explicitly stated.

We thank the reviewer for this insightful comment. We agree that the original Introduction did not sufficiently clarify the specific gaps in GPP sensitivity analysis in China or clearly distinguish the contribution of the present study from previous Noah-MP uncertainty studies.

In the revised Introduction, we added a paragraph explicitly identifying three research gaps: (1) the contributions of different Noah-MP parameterization schemes to GPP simulation uncertainty have not been systematically quantified across China’s diverse ecosystems; (2) the ecosystem- and season-dependent sensitivities of GPP to these parameterized processes remain unclear; and (3) previous studies have rarely combined nationwide sensitivity analysis with site-level GPP evaluation using both flux-tower observations and an independent gridded GPP benchmark product.

We also revised the subsequent objective statement to clarify the distinguishing features of this study. Specifically, this study combines a 48-member Noah-MP multi-physics ensemble, nationwide Sobol’ sensitivity analysis, site-level evaluation using China Flux observations and PML-V2 GPP, and comparisons across vegetation types at seasonal and multi-year scales. These revisions explicitly highlight how the present study extends previous hydrology-focused or limited site-scale carbon-cycle sensitivity studies.

Q2. Line 185, Figure 2b. The MODIS land use data used for this vegetation distribution appears to be a significant source of simulation error. It is recommended to replace the land use distribution map and rerun the simulation.

We thank the reviewer for this important suggestion. We agree that uncertainty in land-cover and vegetation-type representation may propagate into Noah-MP simulations through the assignment of vegetation parameters and plant functional types.

To evaluate the sensitivity of the simulated GPP to the land-cover dataset, we conducted additional Noah-MP simulations using two alternative vegetation maps, namely the HYBMAP dataset (https://doi.org/10.5281/zenodo.10488191) (Luo et al., 2024) and the PFT dataset (https://doi.org/10.5281/zenodo.18448036) (Liu et al., 2026), in addition to the original MODIS land-cover dataset. All three land-cover products were reclassified into the Noah-MP vegetation categories using the same cross-walk procedure, while all other model configurations, forcing data, simulation periods, and parameterization schemes were kept unchanged.

The comparison against the PML GPP product showed highly consistent performance among the three land-cover configurations. The TSS values were 0.5795, 0.5855, and 0.5855 for the MODIS, HYBMAP, and PFT configurations, respectively; all three configurations showed an R value of approximately 0.75 and an RMSE of approximately 0.13 kg m^{-2} mon^{-1}. These results indicate that replacing the MODIS land-cover map resulted in only marginal changes in the overall GPP simulation performance at the national scale.

Therefore, while land-cover uncertainty remains a potential source of local simulation error, the additional experiments suggest that the major systematic biases identified in this study cannot be explained solely by the choice of the MODIS land-cover dataset. The revised manuscript now includes the comparison figure and corresponding discussion in **Section 5.4** and **Figure 1**.

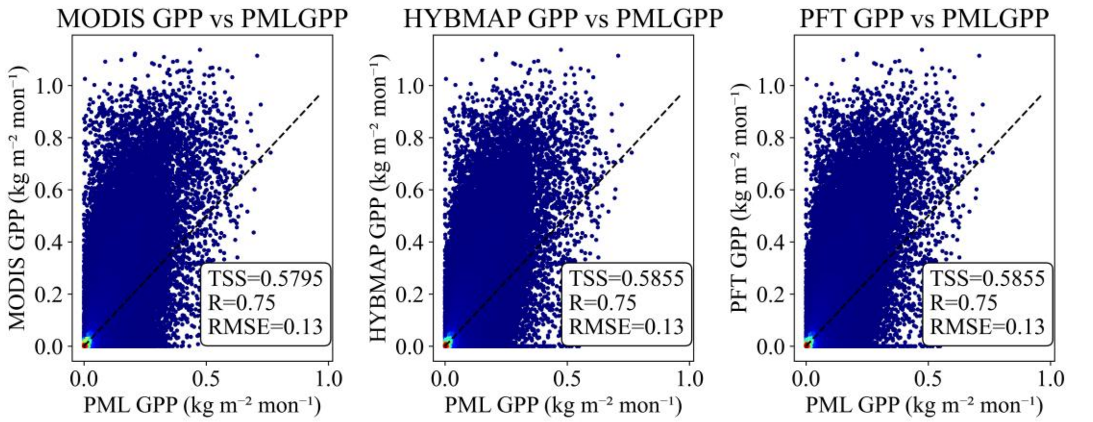

Figure 1. Comparison of monthly Noah-MP GPP simulated using three vegetation-map configurations with PML GPP. Panels (a-c) represent simulations driven by the MODIS, HYBMAP, and PFT vegetation maps, respectively. The dashed line denotes the 1:1 line.

Q3. Line 232. The uncertainty in the forcing data needs to be discussed. For example, although CMFD v2.0 is a solid dataset, it inevitably has uncertainties, especially in station-sparse regions like the western Tibetan Plateau. The paper doesn't address whether/how these uncertainties might propagate through the model and affect the sensitivity results. A brief discussion would strengthen the conclusions.

We thank the reviewer for this important comment. We agree that uncertainties in the CMFD v2.0 forcing data may propagate through the Noah-MP model and affect both GPP simulations and the inferred parameterization sensitivity patterns.

In the revised manuscript, we added a discussion in Section 5.4 (Uncertainties and limitations) to address this issue. We clarify that uncertainties in precipitation, shortwave radiation, temperature, humidity, and wind forcing may affect soil moisture availability, surface energy partitioning, evapotranspiration, stomatal regulation, and ultimately simulated GPP.

We further note that forcing-data biases may influence not only the magnitude of simulated GPP but also the inferred relative importance of the tested parameterizations. For example, biases in precipitation or radiation may shift an ecosystem between water-limited and energy-limited conditions, thereby affecting the apparent sensitivities of the β-factor, radiation-transfer, turbulence, and runoff parameterizations.

We specifically acknowledge that this limitation is particularly relevant in observation-sparse regions, such as the western Tibetan Plateau. Therefore, the sensitivity results in

these regions should be interpreted as conditional on the CMFD v2.0 forcing dataset rather than as absolute estimates of process importance. We also identify multi-forcing ensemble simulations and comparisons with independent meteorological observations as important directions for future work.
The corresponding discussion has been added in Section 5.4.

Q4. Line 298. The 30-times cycle of spin-up may be another source of uncertainty. For conventional meteorological factors, I believe 30 cycles are sufficient for initiation. However, regarding the authors' research objective of GPP, I doubt this will provide enough driving forces for the integration and accumulation of dynamic vegetation mechanisms. I suggest the authors examine the time-varying curves of variables related to vegetation growth (such as aboveground or total dry weight) during the simulation to see if growth and mortality rates have reached a steady state.
We thank the reviewer for this important comment. We agree that the convergence of conventional meteorological and hydrological states alone is insufficient to demonstrate the adequacy of spin-up for simulations involving dynamic vegetation and GPP.
In the revised manuscript, we therefore conducted an additional spin-up convergence assessment for the MODIS-initialized simulation used in the subsequent experiments (Figure 2 and Figure 3). In addition to gross primary productivity (GPP), evapotranspiration (ET), soil moisture (SM), latent heat flux (LH), and sensible heat flux (HFX), we examined vegetation structural, biomass, and carbon-assimilation variables, including leaf area index (LAI), leaf mass (LFMASS), fine-root mass (RTMASS), stem mass (STMASS), net primary productivity (NPP), and total photosynthesis (PSN).
Convergence was assessed over the final 10 spin-up cycles using two indicators: the absolute normalized linear trend and the mean absolute relative change between adjacent cycles. A variable was considered converged when both indicators were below 1.0×10^{-3}. For the carbon, water, and energy variables, normalized slopes ranged from 2.05×10^{-5} to 2.91×10^{-4}, and relative changes ranged from 2.96×10^{-5} to 3.31×10^{-4}, all below the predefined threshold. The vegetation structural variables and biomass/carbon pools also showed negligible variation during the final cycles.
We have added the spin-up convergence method in Section 3.3, a new results subsection in Section 3.3, and the detailed convergence curves in **Figures 2-3**. These analyses confirm that the land-surface states and dynamic vegetation processes reached a quasi-steady state before the GPP simulations and parameter sensitivity analyses were conducted.

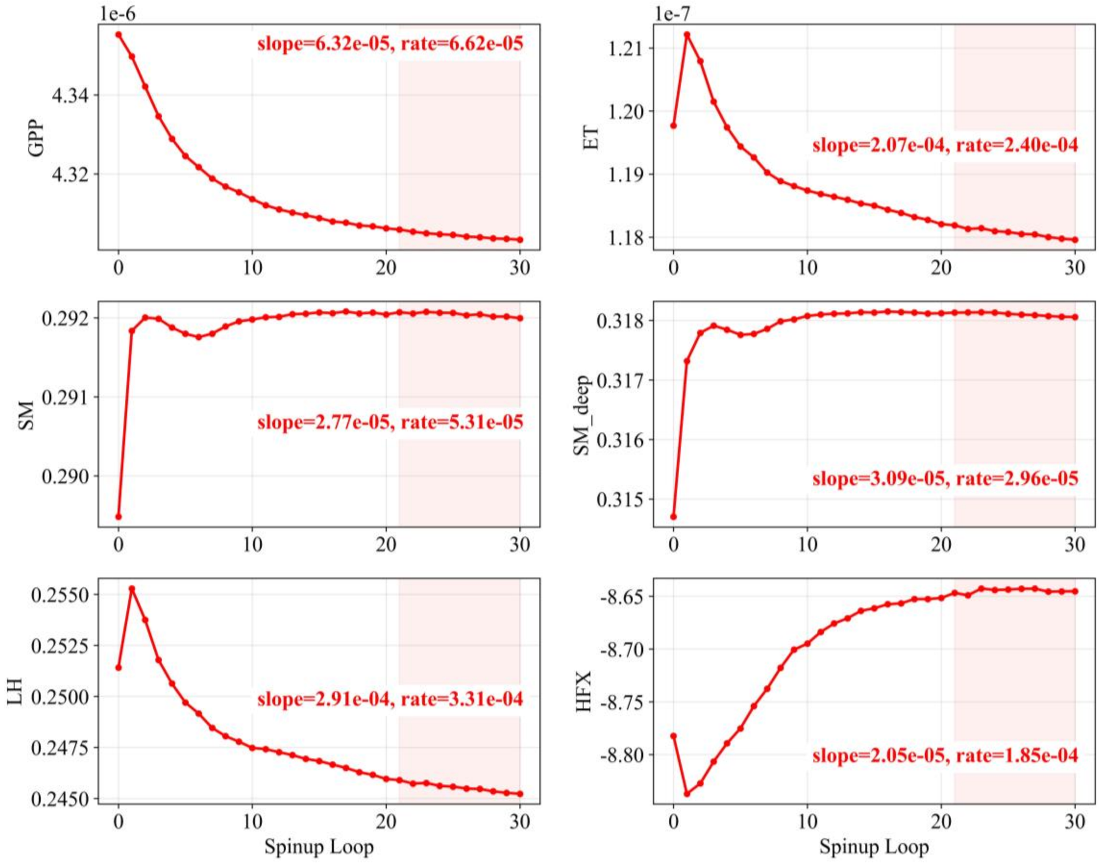

Figure 2. Spin-up convergence of carbon, water, and energy variables for the MODIS-initialized simulation. Each point represents the cycle-mean value for one spin-up loop. Shaded areas indicate the final 10 cycles used for convergence assessment. The annotations show the absolute normalized linear trend and the mean relative inter-cycle change rate. All diagnostic variables satisfied the predefined convergence criterion of 1.0×10^{-3}.

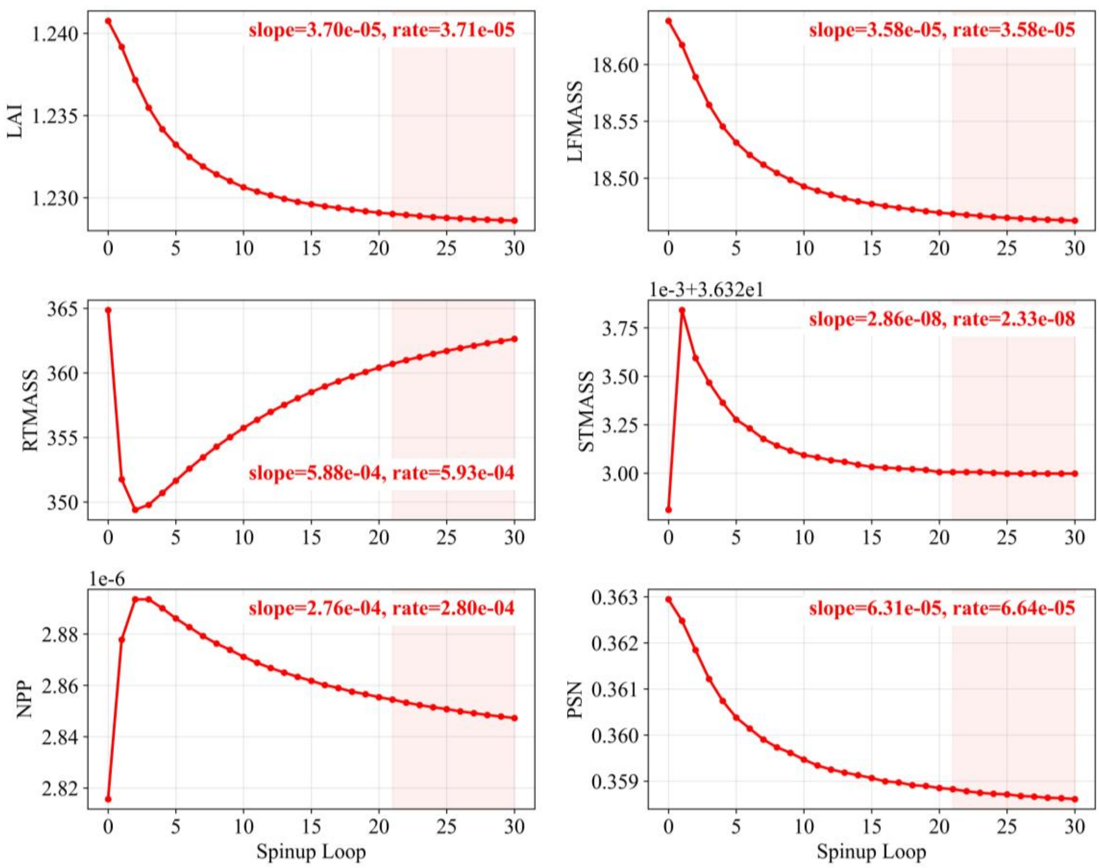

Figure 3. Convergence of vegetation structural variables and biomass/carbon pools during the MODIS-initialized spin-up. Each point represents the cycle-mean value for one spin-up loop. The shaded area denotes the final 10 cycles used to evaluate convergence. Stable LAI, vegetation cover, biomass/carbon pools, and vegetation turnover indicators demonstrate that the dynamic vegetation state reached a quasi-steady condition before the main simulations.

Q5. 4.1.1. It seems the authors' experiments primarily demonstrated reasonably good performance in farmland, with generally poor results in grasslands, and even worse, yielding outcomes like DHF in forests. Given this performance pattern, how did the authors isolate and address this systematic error?

We thank the reviewer for this important comment. We agree that the contrasting performance among cropland, grassland, and forest sites suggests an ecosystem-dependent pattern of model bias. However, the additional diagnostic analysis also reveals substantial heterogeneity among sites within the same ecosystem category. For example, CBF and QYF showed relatively high GPP agreement, whereas DHF exhibited much weaker agreement. Similarly, grassland sites did not show a uniform response: DLG, DXG, and HBG_G01 showed moderate GPP correlations, whereas NMG and XLG showed relatively low correlations.

To isolate this pattern, we conducted an ecosystem-stratified and process-based diagnostic analysis using available flux-tower-derived GPP, APAR, soil water content, latent heat flux, and sensible heat flux. We evaluated both temporal agreement and normalized error magnitude for each variable and further compared the seasonal cycles of these variables at representative forest, grassland, and cropland sites (Figs.4-6). Figure 4 shows the correlation coefficients (R) for each variable across sites; Figure 5 presents the normalized root-mean-square errors (nRMSE); and Figure 6 illustrates the monthly climatological seasonal cycles at representative sites.

The results indicate that low GPP performance cannot be attributed to a single source of uncertainty. At DHF, the low GPP agreement (R=0.40) and high normalized error were accompanied by weak agreement and large errors in APAR and sensible heat flux. The seasonal comparison further showed substantial overestimation of APAR and GPP during the growing season, together with systematic discrepancies in soil water content and surface energy fluxes. This pattern suggests that excessive canopy radiation absorption, biased energy partitioning, and insufficient water-stress limitation may jointly promote unrealistically high carbon assimilation.

The grassland sites showed a different pattern. Although some sites exhibited moderate agreement in individual hydrological or energy variables, their GPP seasonal amplitude or timing remained poorly represented. This finding suggests that grassland GPP errors may also involve phenology, canopy development, radiation interception, and site-specific physiological responses, rather than soil-moisture error alone.

The contrast between DHF and CBF further indicates that vegetation category alone is insufficient to explain the observed GPP bias. Although both sites are represented as mixed forests in Noah-MP, they occur under substantially different hydroclimatic conditions. We therefore interpret the divergent performance as evidence that fixed plant-functional-type parameters and the simplified Ball-Berry stomatal conductance formulation may not fully represent site-specific canopy-atmosphere coupling and water-stress regulation.

Rather than masking these biases through site-specific recalibration, which would compromise the objective of the nationwide multi-physics sensitivity analysis, we addressed them through ecosystem-stratified diagnosis, process-based evaluation, alternative land-cover experiments, and an expanded discussion of the model processes requiring improvement. The corresponding analyses and discussion have been added to Sections 4.1.1 and 5.1.

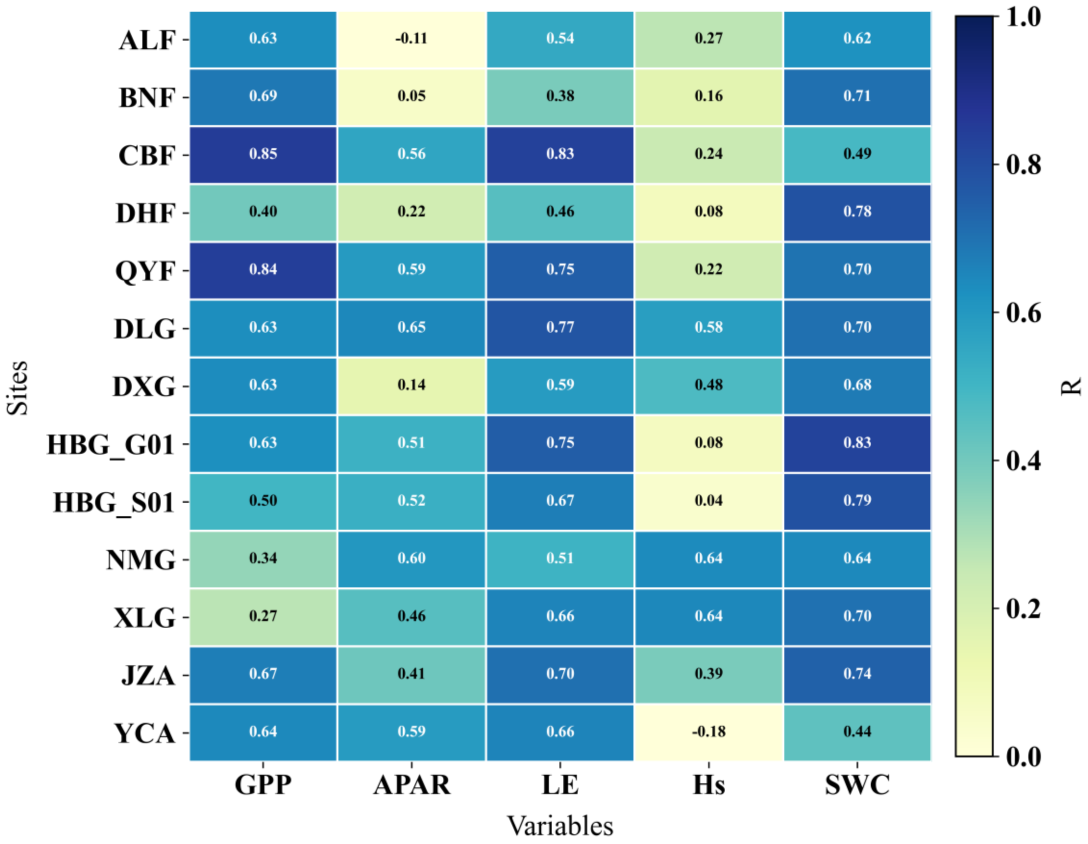

Figure 4. Correlation coefficients (R) for GPP, APAR, latent heat flux (LE), sensible heat flux (Hs), and soil water content (SWC) at flux-tower sites.

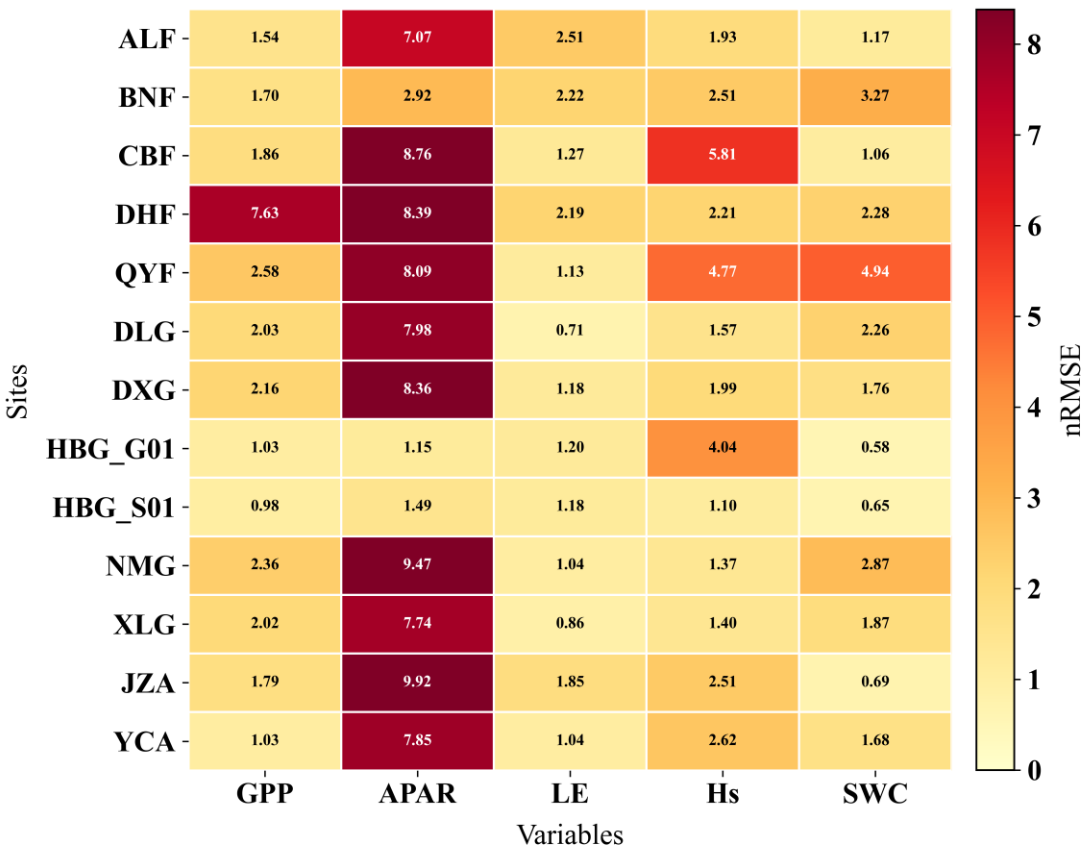

Figure 5. Normalized root-mean-square errors (nRMSE) for GPP, APAR, latent heat flux (LE), sensible heat flux (Hs), and soil water content (SWC).

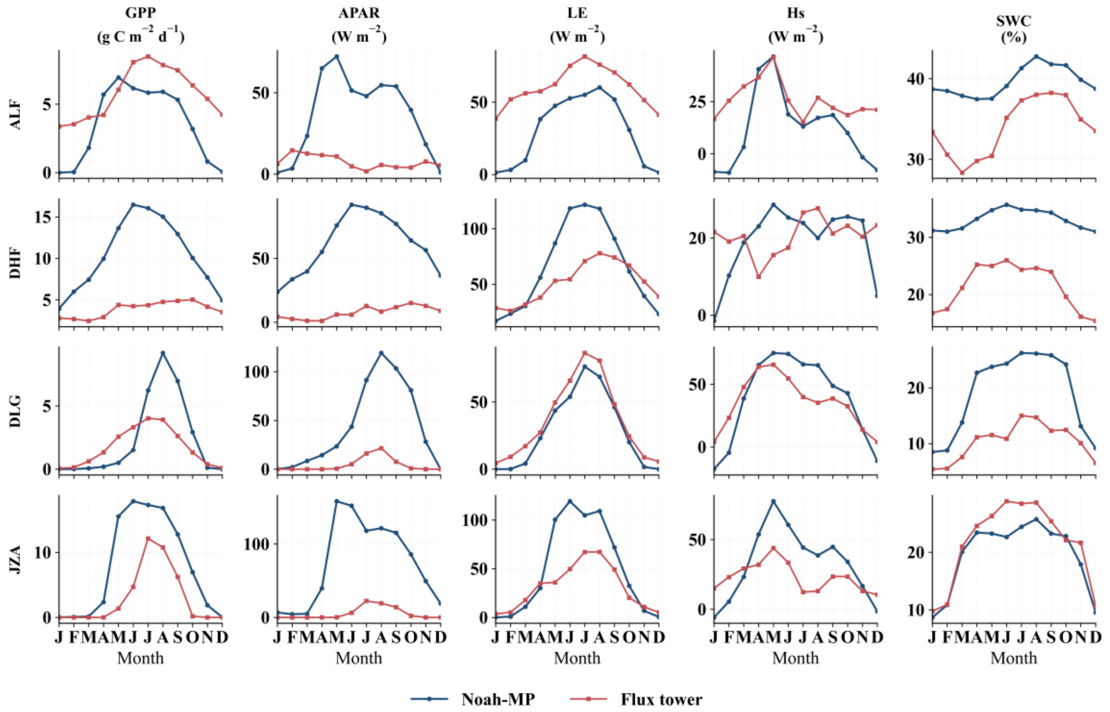

Figure 6. Monthly climatological seasonal cycles of GPP and diagnostic variables at representative flux-tower sites.

Q6. Section 4.1.2. Should all Figure 3 be Figure 4? Meanwhile, the color bars in Figure 4(a-j) are difficult for me to identify. Could the author please create a different plot for the left and right sides? I think the author needs to seriously analyze their results. Especially the qualitative conclusions drawn after line 391; what we see in Figures 4k and l is not simply a matter of slightly higher or lower. The authors' simulations of GPP using Noah-MP showed that, compared to PML, the results were nearly twice or more than twice as high for various land use types, including EBF, DBF, and Cropland, etc. Why?

We sincerely thank the reviewer for the careful examination of this figure and for the constructive suggestions. We corrected the figure citation error in Section 4.1.2, and all incorrect references to Figure 3 have been revised accordingly.

Regarding the figure layout, after reconsideration, we retained the paired-panel configuration of PML and Noah-MP GPP maps. The two datasets are displayed side by side for each season and the multi-year mean using the same color scale, which allows direct comparison of their spatial distributions and seasonal variations. We agree that the original figure could be improved in terms of visual interpretation, and we have revised the figure presentation by improving the color-bar clarity and strengthening the quantitative comparison in the revised manuscript.

We also agree that the differences between Noah-MP and PML are substantial for several vegetation types and should not be described simply as small deviations. The revised manuscript now quantitatively reports that Noah-MP simulated multi-year

mean GPP values were approximately 68.5%, 107.1%, and 106.4% higher than PML estimates for EBF, DBF, and cropland ecosystems, respectively. We further added a process-based discussion indicating that these differences may be related to uncertainties in canopy radiation absorption, soil-water limitation, surface-energy partitioning, and vegetation physiological regulation. The flux-tower-based evaluation further suggests that the differences between Noah-MP and PML are partly consistent with potential overestimation of Noah-MP GPP at several sites.
These revisions have been incorporated into Section 4.1.2 and Section 5.2.

Q7. Line 412. Typo of "sensitivityof". At the same time, I would like to give the author a suggestion in this section(4.2): to conduct regional statistics on these influencing variables according to land use type classification. Now, I can see the constraint relationship between the map and land use type by referring to Figure 2 above. However, the author did not provide a single figure in this entire section, which I think is very irresponsible.
We thank the reviewer for this constructive suggestion and apologize for the insufficient presentation of the quantitative vegetation-type analysis in the original manuscript. We also corrected the typographical error "sensitivityof" to "sensitivity of."
In the revised manuscript, we reorganized Section 4.2 to present both the spatial and vegetation-type-dependent sensitivity patterns of Noah-MP-simulated GPP. Figures 6 and 7 show the spatial distributions of the total Sobol' sensitivity indices and the dominant parameterized processes, respectively. In addition, Figure 8 has been incorporated into this section to provide quantitative vegetation-type statistics of the sensitivity indices for radiation transfer, the β-factor, surface turbulent exchange, and runoff generation.
Specifically, the grid-cell Sobol' sensitivity indices were aggregated according to vegetation-type classes derived from the land-cover dataset. This analysis quantitatively demonstrates that the dominant source of parameterization-related GPP uncertainty differs among ecosystems. For example, radiation transfer is most influential in shrublands, the β-factor is most influential in ENF, savannas, croplands, and grasslands, turbulence shows greater importance in DBF, and runoff generation contributes more strongly in EBF, DNF, and MF ecosystems.
We further clarified that these results represent the relative contributions of parameterized processes to Noah-MP GPP simulation uncertainty, rather than universal ecological controls on actual GPP. The revised section now includes the relevant figures and quantitative vegetation-type statistics, making the connection between spatial sensitivity patterns and vegetation-type classes explicit.
Figure 7. The Sobol' index of the Noah-MP ensemble-simulated multi-year-averaged and seasonal GPP to the four physical processes across different vegetation types. Here, RAD denotes radiation transfer, BTR denotes the β-factor, SFC denotes turbulence, and RUN denotes runoff generation.

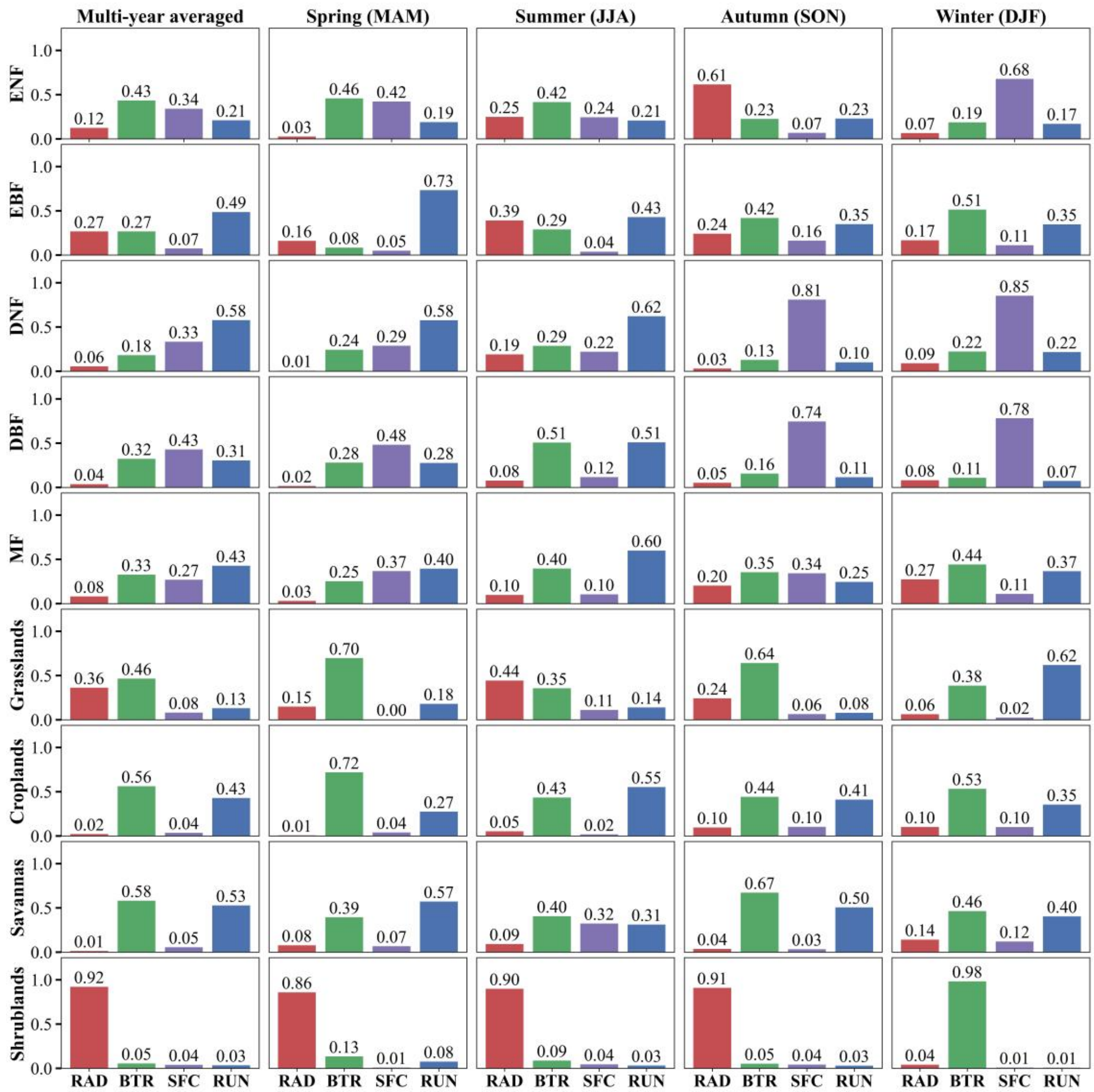

Figure 7. The Sobol' index of the Noah-MP ensemble-simulated multi-year-averaged and seasonal GPP to the four physical processes across different vegetation types. Here, RAD denotes radiation transfer, BTR denotes the β-factor, SFC denotes turbulence, and RUN denotes runoff generation.

Q8. I carefully reviewed the supporting information documents provided by the author. I believe the author needs to examine the significant phase differences between their simulation results and the PML results and explain the reasons for these differences. Furthermore, I suggest that although the goal of this study is GPP, all intermediate diagnostic variables need to be validated in various aspects throughout the simulation process, especially against the flux tower data collected by the author. For example, examine the simulated upward radiation, sensible latent heat flux, soil water, and soil temperature. These can reflect the sources of the author's systematic errors and assist in designing new experiments to improve the GPP simulation level. Only after the overall GPP simulation is reliable should the sensitivity of various parameter combinations be discussed.

We sincerely thank the reviewer for this important and constructive comment. We agree that evaluating GPP alone is insufficient for interpreting model performance and parameterization sensitivity. The sensitivity results should be interpreted in the context of the performance of the associated radiation, water, and energy processes.

In the revised manuscript, we added a site-level process-based diagnostic evaluation using available flux-tower-derived observations of absorbed photosynthetically active radiation (APAR), latent heat flux (LE), sensible heat flux (Hs), and soil water content (SWC) (Figures 4-5). We evaluated the temporal correlation and normalized root-mean-square error (nRMSE) of these variables across sites and further examined their monthly seasonal cycles at representative forest, grassland, and cropland sites.

The new results indicate that GPP errors were not associated with a single common process but instead reflected different combinations of radiation, water, and energy-process errors among sites. For example, at DHF, Noah-MP showed weak agreement for GPP (R=0.40; nRMSE = 7.63) and APAR (R=0.22; nRMSE = 8.39), together with discrepancies in LE, Hs, and SWC. The seasonal comparison further showed substantially overestimated growing-season GPP and APAR at this site. At DLG, the broad seasonal timing of GPP was captured, whereas the simulated amplitudes of GPP and APAR were considerably larger than the tower-derived estimates. At JZA, GPP showed moderate temporal agreement, but APAR and surface-energy variables still exhibited notable disagreement. These results indicate that reasonable agreement in the temporal dynamics of GPP does not necessarily imply accurate representation of the underlying radiation, water, and energy processes.

We also re-examined the seasonal differences between Noah-MP and PML GPP by comparing their monthly GPP cycles at representative sites (Figure 6). We clarified in the revised Discussion that PML is a model-based GPP product rather than a direct observation. Therefore, the phase differences between Noah-MP and PML may reflect differences in the representation of canopy development, radiation absorption, soil-water limitation, surface-energy partitioning, and stomatal regulation in the two modeling frameworks. The comparison with flux-tower-derived GPP further indicates that the disagreement cannot be attributed solely to uncertainty in the PML product, because Noah-MP also showed site-dependent biases in GPP timing and magnitude relative to tower observations.

Direct validation of upward radiation and soil temperature was not conducted because consistent observations were not available across a sufficient number of matched flux-tower sites. This limitation has been explicitly acknowledged in the revised Discussion. We therefore do not claim that all intermediate processes have been fully constrained. Instead, the Sobol' results are interpreted as conditional estimates of parameterization-related uncertainty under the applied model configuration, forcing data, and demonstrated process-level performance.

These revisions have been incorporated into Sections 4.1.1, 5.2, and 5.4.