Response to Reviewer 1 Comments

The authors have addressed most of previous comments, and the manuscript has been substantially improved. Below are a few additional suggestions for consideration:

Response: Thank you very much for your valuable suggestions for revision. Following your advice, we have carefully and thoroughly revised the manuscript. The main revisions are as follows:

- Regarding the framing of the carbon density analysis: We have accepted your suggestion and explicitly framed this work as a methodological contribution in the manuscript—namely, the integration and compilation of existing data—and have revised the wording in the Abstract and Methods sections accordingly.
- Regarding the presentation of the data table: We have clarified the reason for including only the land types (forests and grasslands) that are directly relevant to the "bookkeeping model" methodology employed in this study.
- Regarding the comparative analysis and data consistency: We have made several key updates. First, we have updated the comparison data to the latest Global Carbon Budget 2024 (GCB 2024). Second, we have re-verified our own carbon flux estimates, confirming and explaining the internal consistency between the figure's curve and the table's data regarding the source-to-sink transition point (after 1980). Most importantly, guided by your insightful suggestion, we have completely rewritten the comparative analysis section (Section 4.2) to systematically elucidate the core conceptual differences between various models (bookkeeping models vs. DGVMs).

Point 1. While carbon density is a key component of the study, its novelty may be overstated in the context of a millennium-scale analysis. Given the inherent uncertainties over such extended timescales, the accuracy of current density mapping could be largely obscured. It may be more appropriately framed as a methodological contribution rather than a novel finding.

Response: Thank you for this comment. Revised.

We fully agree with your view and have revised the manuscript accordingly. We have carefully reviewed the entire manuscript and revised the wording in the Abstract (lines 25–27) and Methods section (lines 212–213) to reflect this perspective. Specifically, we now clarify that, based on previously published datasets, our contribution lies in the integration and compilation of existing data.

Point 2. Table 2, how about data for other land types, e.g. cropland, other land? They are not all zeros, correct?

Response: Thank you for this comment.

In the literature and books on carbon density that we collected since 1980, the records primarily cover three major land categories: forest, grassland, and cropland, with only sporadic data available for other land types. For the purposes of this study, we employed the bookkeeping method (a statistical model) proposed by Houghton and Castanho (2023), which incorporates carbon density values for forests and grasslands. Therefore, in Table 2 we present only the data relevant to our analysis, and the carbon density values for other land-use types from the original sources are not displayed here.

Point 3. It is becoming clear that over the past few decades, LUC in China is shifting from a net CO2 source to a net sink (or at least neutral). This is evidenced by the latest GCB 2024 (Fig. 7b in Friedlingstein 2024; https://doi.org/10.5194/essd-17-965-2025) and a recent Nature Climate Change study (Fig. 1 and 3a in Zhu 2024; http://dx.doi.org/10.1038/s41558-025-02296-z). In particular, the Houghton model (H&C) in GCB 2024 shows negative fluxes for China since the 1960s (data is here: https://globalcarbonbudgetdata.org/latest-data.html). While your Fig. 10 seems to show negative fluxes since 1970-80s for "this study", Table 3 presents a positive 2.25 Pg C for 1980-2019. Any clarification on this discrepancy? Additionally, why do you choose older GCBs instead of the latest one? Please clarify or justify. I would encourage an updated discussion on "4.2 Comparison with previous estimates”. In the discussions, please also note that the concept of LUC emissions differs between bookkeeping vs. DGVM models.

Response: Thank you for this comment. Revised.

Section 4.2 (see lines 545-631) presents a completely rewritten comparative analysis against previous research. Its central focus is the systematic clarification of conceptual distinctions between different estimation approaches. The analysis categorizes existing estimates into three main types—Bookkeeping Models (BKM), Dynamic Global Vegetation Models (DGVM), and National Reports—and explains the fundamental differences in their respective accounting scopes. Please note that the previous Figure 10 is now referred to as Figure 11 in this revised document because of structural changes and the insertion of new figures.

(1) Update to Figure 11. We have updated Figure 11 by replacing the comparison dataset “GCB 2019” with the latest “GCB 2024.” In addition, we have incorporated the carbon emission data for China (1981–2020) from Zhu et al. (2025) and added corresponding descriptions in the text (see lines 545-631).

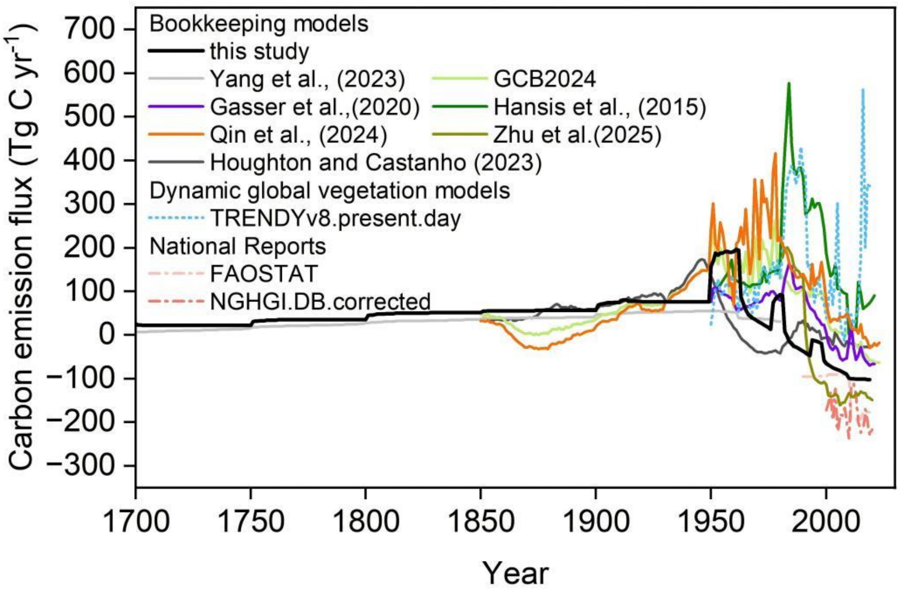

Figure 11. Chinese historical land-use change-induced carbon emission flux estimated by different methods. For detailed descriptions of all data sources, models, and methodologies shown, please refer to the notes of Table 3.

(2) Verification of carbon flux estimates. We have re-checked the carbon flux estimates in this study, which are also available on Zenodo

(https://doi.org/10.5281/zenodo.14557386, 2025). Our results confirm that the carbon flux becomes negative (indicating a carbon sink) after 1980, rather than after 1970. Therefore, the curve shown in Figure 11 is not in conflict. The possible confusion may be due to the long time span of the curve (1700–2020) in Figure 11, which makes it difficult to visually distinguish between 1970 and 1980. You may refer to our openly available dataset for the exact values.

(3) Clarification of LUC emission concepts. In response to your suggestion to note that the concept of LUC emissions differs between bookkeeping models and DGVMs, we have emphasized this distinction in Section 4.2 (see lines 545-631).

We sincerely thank you for your two rounds of careful and professional reviews, which have substantially improved the quality and scientific rigor of our manuscript.

Response to Reviewer 2 Comments

Thank you for addressing most of my previous concerns. However, my main concerns in this second round focus on land-use transition matrix construction, Figure 10, and Table 3, which require substantial improvements for methodological validation and systematic comparison. The current approach lacks rigor in several key areas that undermine the study's credibility.

Response: We sincerely thank you for your profound and constructive feedback during this round of review.

In response to your key concerns regarding the construction of the land-use transition matrix and the associated uncertainties, and the incomplete comparison of different studies (Table 3, Figure 10, now Figure 11 in the revised manuscript), we have undertaken comprehensive, in-depth revisions to enhance the scientific rigor of our study.

The revisions are centered around three aspects:

(1) Regarding the selection of allocation rule in land conversion, following the review's suggestion, we additionally included the area-weighted method. Instead of using a 'probabilistic' approach, we further included the 'forest-priority' method, which, together with the 'grassland-priority' method used in the original manuscript, formed two opposing land conversions and helped building a full range of possible carbon fluxes.

In particular, we apologize for the fact that this comment was raised by the reviewer in the previous round of comment but not fully addressed. We now made our best effort to address this point.

(2) Inspired by the review's comment, we now improved the uncertainty quantification in the revised manuscript by incorporating the uncertainty caused by three sources: the reconstructed land-use data used as model input, the land-use transition rule, and the carbon density parameters applied in carbon flux accounting. In particular, the uncertainty caused by the allocation is a new aspect that is inspired by the reviewer.

(3) We improved the consistency, clarity and completeness of Figure 11 and Table 3. More importantly, the Section 4.2 was substantially revised to accommodate a complete, logic-based comparison among different studies. This section now begins by addressing the fundamental conceptual differences in the core concept of “land-use change emissions” among various estimation methods before proceeding to the comparison of different datasets. This restructuring provides a clearer logical flow and a more rigorous comparison among different data sources.

We hope that these revisions have adequately addressed your concerns and strengthened the methodological foundation and credibility of our conclusions. Please refer to our detailed responses to each of your comments below.

Thank you once again for your valuable guidance.

Point 1. Land-use transition matrix construction lacks methodological validation

The priority-based rules for allocating land-use transitions from aggregate area data introduce substantial uncertainty that remains unaddressed. The authors' justification that forest-to-cropland and forest-to-other-land have identical response curves misses the point—the issue is not the final carbon calculation but the arbitrary nature of the allocation rules themselves. Without testing alternative methods (area-weighted, probabilistic) or demonstrating robustness across different rule sets, the reliability of the entire reconstruction is questionable. This methodological uncertainty compounds over the millennium timescale and undermines confidence in the results.

Response: Thank you for your valuable feedback regarding the allocation rules. We fully agree on your suggestion that the choice of the allocation rule in land conversion can cause uncertainty in the estimated carbon fluxes, and therefore, alternative allocations rules should be explored.

Following your kind suggestion, to systematically assess the uncertainty introduced by the allocation method, we have constructed and compared multiple allocation rules and, based on this, further developed a more comprehensive uncertainty assessment framework. The revision details are as follows:

(1) Setup and comparison of allocation rules (For detailed revisions, please see lines 320-346 of the main text, where all changes have been highlighted in red). To comprehensively assess the impact of different allocation rules on the results, we designed and compared the following three logically distinct rules.

The first rule remains the grassland-priority rule that was used in our original manuscript, which, to the best of our knowledge, has the highest accordance with the reality in China’s land use history. More importantly, this rule maintains the internal consistency with the reconstruction methodology of the historical land-use dataset used in this study.

Following the review’s suggestion, we used the area-weighted allocation method as the second rule. This rule assumes that outgoing land conversions are allocated to different incoming land types in proportion to their area's share of the total outgoing area.

For the third rule, instead of using a probabilistic-based approach as suggested by the reviewer, we used a forest-priority method which is in contrast to the ‘grassland-priority’ method. We argue that this method, together with the ‘grassland-priority’ method, helps to provide a minimum-maximum boundary in land use change areas, which further provides a complete estimate of the uncertainty in the derived land use carbon fluxes.

The details of these three allocation rules, along with a concrete example, are described in the revised manuscript of lines 320-346.

(2) Uncertainty assessment framework (For detailed revisions, please see lines 370-390 of the main text, where all changes have been highlighted in red).

The original uncertainty assessment incorporates only the uncertainties caused by land use data and the carbon density data but not the uncertainty from the allocation rule in the land conversion. The review’s comment reminds us that the allocation rule is a critical source of uncertainty that we failed to account for in our original manuscript.

Hence, following the review's suggestion, we improved the uncertainty quantification in the revised manuscript by incorporating the uncertainty caused by three sources: the reconstructed land-use data used as model input, the land-use transition rule, and the carbon density parameters applied in carbon flux accounting. More specifically, for each of the three allocation rules, we performed a Monte Carlo simulation consisting of 1000 iterations to account for the uncertainties in the land use data and the carbon density parameters. To establish the final uncertainty in the estimated carbon flux, we aggregated the annual carbon emissions from all 3000 simulations (3 allocations rules × 1000 iterations for each rule) for each year and selected their maximum and minimum values.

We thank the review's comment which encourages us for a more complete estimation of the uncertainty as described above. The details are described in the lines 370-390 of the revised manuscript.

(3) Uncertainty and sensitivity analysis results (For detailed revisions, please see lines 480-516 of the main text, where all changes have been highlighted in red).

Following the reviewer's suggestion, we have constructed a more comprehensive uncertainty assessment framework. In this section, we present and interpret the results from this comprehensive uncertainty analysis.

To examine the influence of different allocation rules, we compared the annual land-use transition matrices generated by the grassland-priority method and the area-weighted method (see Appendix Fig. C2). The results indicate that although certain transition types exhibit numerical differences in area, the primary transition processes (e.g., the conversion of forest to cropland) show a high degree of consistency. This demonstrates that the core conclusions of our study are robust.

Nevertheless, the choice of allocation rule remains a significant source of overall uncertainty. The results of the comprehensive uncertainty assessment are presented in the revised Figure 10. This figure displays the millennial-scale carbon emission flux estimated using the grassland-priority allocation rule (black line), along with a comprehensive uncertainty interval (gray shaded area). This interval was constructed by systematically integrating the results from all 3,000 Monte Carlo simulations (1000 for each of the three allocation rules) and selecting their maximum and minimum values. It thereby comprehensively reflects the combined uncertainties from the three main sources: input data, the allocation rule, and carbon density parameters.

Furthermore, a period-based analysis of the uncertainty (see Appendix Table C1) reveals its temporal evolution. The analysis shows that after 1982, when land use shifted to a carbon sink, the uncertainty exhibits a distinct asymmetry: the downward uncertainty (58.91 Tg C) is significantly larger than the upward uncertainty (34.56 Tg C). This result indicates that the actual sink strength of China's land use during the modern observational period is likely stronger than our current best estimate.

In summary, the new comprehensive assessment framework proposed in this study, by systematically considering multiple sources of uncertainty, provides a robust quantification for the estimated carbon fluxes. This analysis reaffirms the trends and key turning points in China’s millennial-scale land-use carbon budget.

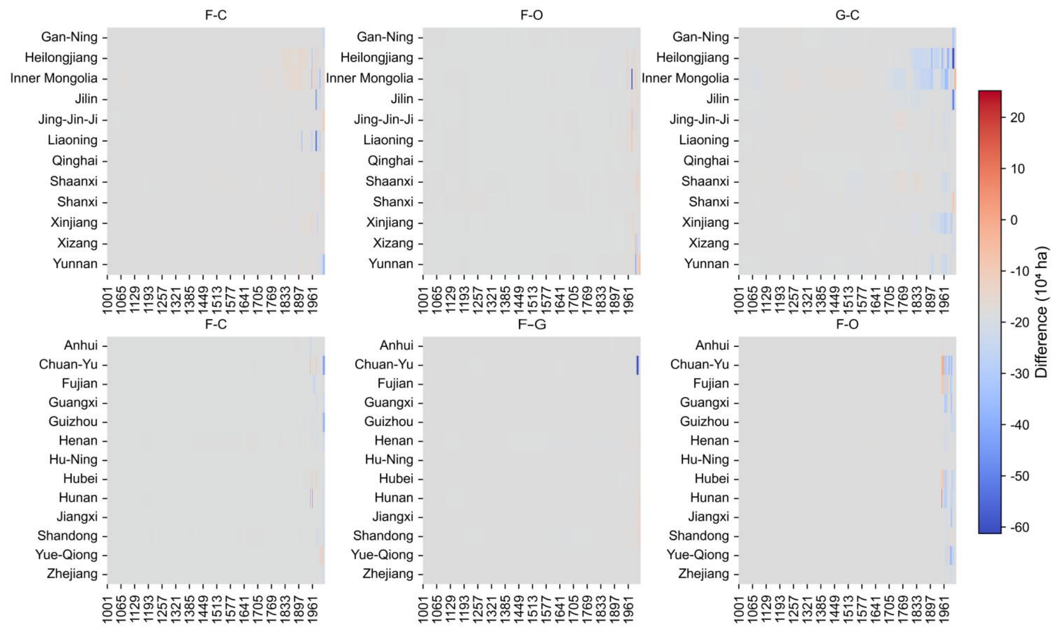

Figure C2. Differences in annual land-use transitions between the grassland-priority and area-weighted allocation methods. F-C denotes the conversion between Forest and Cropland, F-G represents Forest-Grassland conversion, F-O represents Forest-Other land conversion, and G-C represents Grassland-Cropland conversion. A positive difference indicates that the grassland-priority result is lower than the area-weighted result, and vice versa.

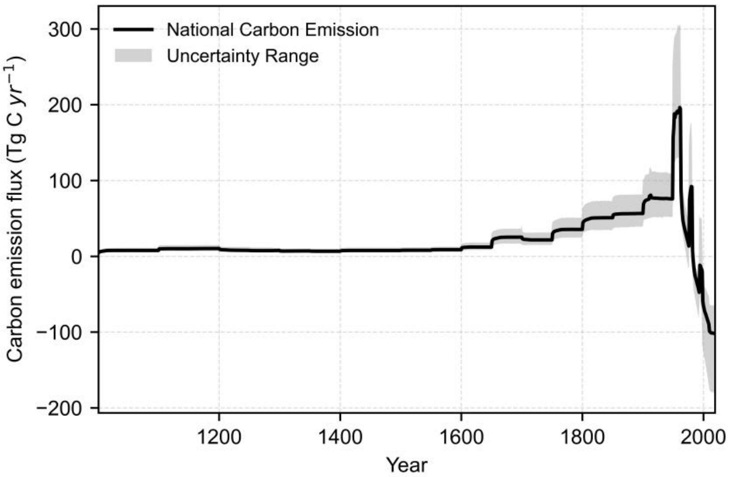

Figure 10. Carbon emission fluxes from land-use change in China with their uncertainties. The black line represents the mean result of 1000 Monte Carlo simulations using the grassland-priority allocation rule in land use conversion. The gray shaded area represents the uncertainty interval defined by the maximum and minimum values across all 3000 simulation runs (1000 iterations for each of the three allocation rules: grassland-priority, area-weighted, and forest-priority). This interval incorporates the combined uncertainties from input data, carbon density parameters, and the allocation rule. For details, refer to the sections 2.3.2 and 2.3.3.

Table C1. Carbon emissions from land-use change and their uncertainties in different historical periods

Period	Mean annual estimate (Tg C)	Upward uncertainty	Downward uncertainty
Pre-1900	16.35	7.49	5.20
1900-1949	75.51	37.31	24.65
1950-1982	102.95	74.35	44.48
Post-1982	-60.90	34.56	58.91

Notes: All values are in the unit of Tg C/yr. The mean upward uncertainty is the average of the annual differences between the maximum value and the estimate within each period. The mean downward uncertainty is the average of the annual differences between the estimate and the minimum value.

Point 2. Figure 10 legend and data sources lack essential details. -Data sources for all model results need clear citation.-Update to latest GCB2024 data instead of GCB2019. Figure 10 presents Gasser (2020), Hansis et al.(2015), and Houghton (2023) alongside GCB2019 without explaining that these three studies are the component models underlying GCB estimates. This may mislead readers about the independence of these approaches. -NGHGI.DB and NGHGI.DB.corrected are undefined. What specific corrections were applied and how? -Justify TRENDYv8 selection (if due to additional scenarios isolating LASC effects, state this explicitly)

Response: Thank you for these valuable comments. In response, we have comprehensively revised and restructured Section 4.2 Comparison with previous estimates to improve clarity regarding data sources, definitions, and their interrelationships. Specifically, we have implemented the following key changes:In response to your specific points in Comment 2 regarding data sources, definitions, and their logical relationships, we have systematically rewritten and comprehensively reorganized Section 4.2 Comparison with previous estimates.

(1) Updating to GCB2024: As suggested, we have updated our comparison dataset from GCB2019 to the latest GCB2024 version, ensuring our analysis is based on the most current data.

(2) Clarifying data relationships: We have revised the text, and the footnotes of Table 3 to explicitly state that the GCB2024 estimate is the mean of its component bookkeeping models. This directly addresses the potential for misunderstanding the independence of these datasets.

(3) Defining datasets and justifying selections: We have added clear definitions and justifications for our choice of comparison datasets:

NGHGI.DB.corrected is now defined as the original national inventory data adjusted by subtracting carbon fluxes from indirect environmental changes (e.g., CO_2 fertilization). This correction conceptually aligns its accounting boundary with our bookkeeping model.

The selection of the TRENDYv8 present.day scenario is now justified in the text. This scenario is specifically designed to be conceptually consistent with bookkeeping models by using fixed, modern environmental conditions.

(4) Improving the section's logical structure: Beyond addressing these specific points, we have fundamentally restructured Section 4.2. As part of this effort, we have substantially revised Table 3 to serve as a comprehensive summary for the numerical comparison, now supported by detailed footnotes that provide all necessary definitions. The main text now precedes this comparison by first explaining the core conceptual differences between the three main estimation methods (Bookkeeping

Models, DGVMs, and National Reports). This provides a clearer framework for the reader to understand the discrepancies between different estimates.

Table 3. Comparison of average annual carbon flux estimates caused by land-use change in China

Reference	Model Type[1]	Average annual carbon flux (Tg C yr^{-1})			
		Pre-1900	1900-1949	1950-1980	Post-1980
This Study	BKM	40.50	75.51	106.39	-55.23
Yang et al. (2023)	BKM	25.79	51.31	42.76	/
Ge et al. (2008)	BKM	18.48	49.70	/	/
Yang et al. (2019)	BKM	23.54	42.25	149.36	/
Houghton and Castanho (2023)	BKM	48.14	100.59	6.57	-6.98
Qin et al. (2024)	BKM	-3.81	77.06	236.13	73.68
Gasser et al. (2020)	BKM	/	/	82.50	20.23
Hansis et al. (2015)	BKM	/	/	126.70	212.73
GCB2024[2] (Friedlingstein et al., 2025)	BKM	21.20	83.64	164.33	36.27
Zhu et al. (2025)[3]	Stock-Difference Method	/	/	/	-49.33
TRENDYv8.present.day[4,5]	DGVM	/	/	107.40	190.39
FAOSTAT[4,6]	National Reports	/	/	/	-121.06
NGHGI.DB.corrected[4,7]	National Reports	/	/	/	-180.36

Note: The values in the table represent the average annual carbon fluxes (Tg C yr^{-1}) for the specified time periods. A "/" indicates that data for that period is not available or not provided in the source study.

[1] Model Type: BKM (Bookkeeping Models) aim to estimate carbon fluxes from direct anthropogenic land-use activities; DGVM (Dynamic Global Vegetation Models) simulate the integrated response of ecosystems to both land use and environmental changes, which includes the additional loss of sink capacity (LASC) that would otherwise occur, for example, on an actually cleared forest (refer to Gasser et al., 2020 for a detailed explanation); National Reports (NGHGI) are based on the IPCC's "managed land proxy" principle for accounting, which includes carbon sink driven by both direct anthropogenic land use and also environmental effects, but not the LASC.

[2] The value for GCB2024 is the mean of the carbon fluxes derived by the four bookkeeping models: Qin et al. (2024), Houghton and Castanho (2023), Gasser et al. (2020), and Hansis et al. (2015).

[3] Zhu et al. (2025) employs a stock-difference method, constructing high-resolution, dynamic carbon stock maps by integrating remote sensing, inventory data, and machine learning, and calculates the flux from changes in these maps over time.

[4] Data for TRENDYv8.present.day, NGHGI.DB.corrected, and FAOSTAT are from Obermeier et al., (2024).

[5] TRENDYv8.present.day represents DGVM simulation results run under fixed, modern environmental conditions (climate and CO_2 concentration), designed for conceptual alignment with bookkeeping models that use modern carbon densities.

[6] FAOSTAT provides bottom-up estimates by applying IPCC guidelines to country-reported activity data (e.g., from the Forest Resources Assessment) and geospatial information. Conceptually, these estimates are closer to bookkeeping models as they often do not include the indirect environmental effects (such as CO_2 fertilization).

[7] NGHGI.DB.corrected is derived from the original national inventory data (NGHGI.DB) by subtracting the carbon fluxes on "managed land" that are caused by indirect environmental changes (e.g., CO_2 fertilization), as estimated by DGVMs. This makes this estimated conceptually aligned with the bookkeeping models.

Point 3. Table 3 is incomplete and inconsistent with Figure 10. The current Table 3 lacks systematic collection of comparable studies and shows inconsistency with Figure 10 content. While Figure 10 includes multiple DGVM studies for China, Table 3 only presents Yu et al. as the sole DGVM representative without justification for this selective inclusion, and other bookkeeping model results as well as NGHGI/FAOSTAT shown in Figure 10 are not included in Table 3.

Response: Thank you for highlighting this critical inconsistency.

We agree that the original Table 3 was incomplete and not fully aligned with the figure. In response, we have comprehensively revised Section 4.2, with a particular focus on reconstructing Table 3 to ensure it is systematic, comprehensive, and consistent with Figure 11 (previously Figure 10). For detailed revisions, please see lines 553-577 of the main text, where all changes have been highlighted in red. The key revisions are as follows:

(1) Reconstructing Table 3 for Consistency and Clarity: The original Table 3 has been completely replaced. The new version now systematically includes all comparable studies presented in Figure 11, resolving the inconsistency you pointed out. To facilitate a more standardized comparison, the table is now structured to show period-based annual average fluxes for each study, providing a much clearer quantitative summary.

(2) Clarifying the Scope of the Table and Figure: We have clarified why minor differences in content between the table and figure persist. The revised table notes now explicitly explain that Table 3 includes some studies not plotted in Figure 11.

This is because those sources only report cumulative fluxes over a period and lack the annual time-series data required for plotting.

(3) Optimizing the Underlying Comparison Framework: We recognized that the inconsistency you identified was symptomatic of a weaker analytical framework in the original manuscript. Therefore, we have rewritten the entire Section 4.2. The new structure first establishes the conceptual differences between estimation methods (Bookkeeping Models, DGVMs, National Reports) before presenting the numerical comparison in the revised Table 3 and Figure 11. This provides a much more rigorous and logical foundation for comparing the different estimates.

Point 4. The statement "estimates in this study fall within the range of existing model estimates at an intermediate level" is incorrect. Figure 10 shows your results are among the most negative values post-2000 (excluding NGHGI data due to different definitional boundaries).

Response: Thank you for your keen observation; you are absolutely correct.

Our previous generalization that our estimate was at an "intermediate level" was inaccurate, and we have thoroughly reflected on this and made a complete correction. The key revisions are as follows:

(1) Removal of the inaccurate statement: We have completely removed this inaccurate generalization.

(2) Providing a More Rigorous Positioning and Discussion: The issue you identified stemmed from a less rigorous comparison framework in our original manuscript. As mentioned in our responses to your previous points, we have rewritten the entire Section 4.2, establishing a rigorous "concepts first, data second" comparison framework.

Within this new framework, instead of making a broad generalization, we use specific data comparisons to reach a more precise conclusion that aligns with your observation: Our results clearly indicate that China's land use has been a significant carbon sink (-55.23 Tg C yr^{-1}) in recent decades. This finding not only stands in stark contrast to mainstream models that rely on global-scale datasets and show a carbon source (e.g., GCB2024), but is also corroborated by the results of another study using localized data (Zhu et al., 2025), together revealing a strong carbon sink signal that has been missed by global models.

This more accurate positioning and discussion are now presented in detail in the revised Section 4.2 (lines 545-631). We thank you again for your correction, which has led to a more precise and insightful interpretation of our own findings.

Response to Reviewer 3 Comments

General Comments. This paper estimates carbon emissions from land-use change in China over the past millennium, which is highly relevant and aligns well with ESSD's scope. The use of historically reconstructed land-use datasets—based on China's unique archival records rather than proxy indicators like population—is particularly valuable. While the authors have addressed some concerns in prior revisions, several issues remain regarding paper structure, integration of historical and modern land-use data, spatial resolution of historical datasets, and practical applications of the results. Additionally, some of my comments overlap with previous reviewers' feedback; I urge the authors to prioritize these shared concerns.

Response: Thank you for your insightful and constructive comments, which have been crucial for improving our manuscript. We have thoroughly revised the paper based on your suggestions, focusing on three key areas: strengthening our scientific rationale, improving the logical structure, and clarifying the practical applications of our dataset.

Key revisions include restructuring the manuscript for a clearer separation of the Methods, Results, and Discussion sections. We have also expanded our methodology to better explain the integration of historical and modern data and have introduced a new sensitivity analysis to quantify the impact of our core assumption of static carbon densities. Finally, to highlight the dataset's value as encouraged by ESSD, we added a dedicated section on its applications in climate research and policy assessment.

We believe these changes substantially strengthen the paper. Below are our detailed responses to each of your points. Thank you again for your time and expertise.

Specific Comments.

Point 1. -Data Integration Issues. Clarify how reconstructed data (e.g., cropland from tax records) align with survey-based statistics (e.g., the Second and Third National Land Surveys). Land-use definitions evolved between surveys (e.g., the Second

National Land Survey [2009] and Third National Land Survey [2019]). Discuss potential errors introduced by these definitional shifts.

Response: Thank you for your valuable feedback. Revised.

We have carefully considered the issue of inconsistent statistical calibers in our data integration. Your point is crucial. In processing the land-use data from 1980 to the present, we faced two primary options: one being the annual-resolution land use/cover datasets derived from remote sensing interpretation, and the other being the national-level, survey-based statistical data, namely the Second and Third National Land Surveys. We chose the latter primarily because the reconstructed historical data we used is more closely aligned and compatible with the national survey data in terms of its sources, methodologies, and nature (e.g., statistics and mapping based on administrative units). We believe that linking datasets of a similar nature helps maintain consistency in the long-term trends and mechanisms.

We fully agree with your observation that even between the Second and Third National Land Surveys, the land classification standards have evolved and differ, which poses challenges for direct data linkage. To minimize the errors introduced by these definitional discrepancies, we performed specific harmonization and adjustments for the most sensitive land class in our carbon flux model: forest. Specifically, the original research literature for the historical forest data explicitly states that its definition of ‘forest’ is conceptually closest to the ‘closed forest land’ sub-category in current classification standards. Therefore, when linking with modern data, we did not use the total area of the primary ‘forest land’ category. Instead, we precisely extracted the data for the ‘closed forest land’ sub-category from both the Second and Third surveys to ensure maximum definitional consistency with the historical reconstructed data.

Despite these efforts, we acknowledge that the definitional evolution of other land classes (e.g., cropland, grassland) across different survey periods, along with the inherent discrepancies in statistical calibers between the reconstructed and survey-based data, remains a source of uncertainty in this study. These differences will inevitably affect the accuracy of the final carbon budget estimation. We will explicitly address this point in the discussion section of our paper and identify it as an important area for future research, which could be advanced through data fusion or the development of more optimal classification conversion algorithms (see lines 645-655 of the manuscript).

Point 2. -Line 141. are considered highly credible. Cite references for this statement.

Response: Thank you for this comment. Revised.

Thank you for your valuable feedback. Regarding your comment on the supporting evidence for our statement that the National Land Survey data "are considered highly credible" (Line 141), we have carefully considered the point and revised the manuscript. We fully agree that providing a clear justification for the reliability of this key dataset is essential. In our revision, we reflected on the best way to establish this credibility. For official census data of this nature—organized by the highest state administrative body and mobilizing national resources—its authority and reliability are typically accepted as a consensus or a benchmark in the academic community. It serves as a foundational starting point for research, rather than a debatable claim requiring repeated justification. Therefore, we concluded that the most rigorous and direct method to demonstrate its credibility is not by citing an indirect evaluation from another study, but by elucidating the rigorous nature of the data production process itself. Based on this reasoning, we have revised the original, more general statement and replaced it with a specific description of the survey process. Please see the revised text in lines 142-143.

Point 3. -Line 145. The text describes vegetation carbon density first but later details soil carbon density before vegetation. Revise for logical flow.

Response: Thank you for this comment. Revised. (Line 212-213)

Point 4. -Temporal Stability of Carbon Densities. Soil/vegetation carbon densities are treated as static over the millennium. The Discussion notes this limitation, but quantify its impact: Would assuming stable densities overestimate or underestimate emissions?

Response: Thank you for your insightful suggestion to quantify the impact of assuming static carbon density over time. Revised.

We fully agree that this is a critical scientific issue. Accordingly, we have designed and completed a sensitivity analysis to assess the potential effects of this assumption on our estimation results. Our analysis is based on the posited systematic differences between historical and modern carbon pools in vegetation and soil. We hypothesized that historical vegetation carbon density was likely systematically lower than modern levels, a premise primarily based on the limited 'CO_2 fertilization effect' under significantly lower pre-industrial atmospheric CO_2 concentrations (approx. 280 ppm vs. >420 ppm today). Conversely, we posited that historical soil carbon density was likely higher than the modern average, mainly due to less intensive anthropogenic disturbance, which allowed soil organic carbon pools in extensive ecosystems to remain closer to a state of natural saturation.

Based on this rationale, we designed a scenario assuming that historical vegetation carbon density was 20% lower and soil carbon density was 20% higher than modern values. After recalculating based on this scenario, we conducted an in-depth analysis of the annual differences between the new and original estimates, revealing distinct temporal patterns. During the carbon source periods, which constitute the vast majority of the study period (approx. 982 years), the new estimates were consistently lower than the original values, with a mean annual difference of approximately -2 Tg/yr, indicating a smaller and more stable range of deviation. In contrast, during the few years identified as carbon sink periods (approx. 37 years), influenced by the intense land-use change during those times, the discrepancy between the two estimates showed greater uncertainty and volatility, with differences ranging from -5 to +11 Tg/yr. This period-segmented analysis indicates that our original methodology may lead to a systematic overestimation of carbon fluxes, and that the uncertainty of this estimation is particularly pronounced during carbon sink periods. We believe this new, more in-depth analysis substantively addresses your concerns and significantly enhances the rigor of our paper's discussion on uncertainties (the detailed revisions in the main text can be found in lines 676-685).

Thank you again for your valuable feedback, which has greatly improved the quality of our research.

Point 5. -Line 164–165. Briefly summarize the framework of the transfer function for bulk density estimation. Technical details can remain in cited sources.

Response: Thank you very much for your valuable feedback. Revised.

In our study, for sample points that lacked measured bulk density data, we employed an empirical transfer function established and validated in Yang et al. (2007) for estimation. This function is based on the significant negative correlation between soil organic matter (SOM) content and bulk density, a relationship that has been widely confirmed in soil science studies. According to that paper, the specific formula for estimation is:

$$BD = 0.29 + 1.2033 \times e^{-0.0775 \times SOM}$$

where BD is the bulk density to be estimated (unit: g/cm^3) and SOM is the percentage of organic matter content in the corresponding soil layer (%). The model's goodness-of-fit (r^2) is 0.81 ($p<0.01$), which indicates a high degree of reliability. To make our research methods clearer and more transparent, we have followed your suggestion and added this specific formula and explanation to the methods section of the revised manuscript. Please see lines 232-235 for the detailed revisions in the main text.

Point 6. -Table 2. Add province/region codes (e.g., "No.1" for Jing-Jin-Ji) to align with Figure 1.

Response: Thank you for this comment. Revised.

Following your recommendation, we have revised Table 2 to ensure its consistency with Figure 1. We have added a new column, "Code," and reordered the rows to align with the regional numbering (No. 1, No. 2, ...) presented in Figure 1. This modification greatly improves the coherence between the table and the figure. The revision can be found on lines 247-250.

Point 7. -Figure 2. Explain color schemes in the flowchart. In other words, What is the meaning of each color in the flowchart?

Response: We thank the reviewer for their constructive feedback and agree that the color scheme in Figure 2 required clarification. Revised.

To address this, we have added an explanatory sentence to the figure caption, defining the module represented by each color (please see lines 258-260 of the revised manuscript). We are confident that this revision improves the clarity of our research framework and the overall readability of the figure.

Point 8. -Line 195. Spatially explicit cropland/forest/grassland data exist (e.g., SCES literature). Justify why provincial-scale aggregation was used instead.

Response: Thank you for your valuable feedback. The question you raised regarding our choice to use provincial-scale summary data is a critical methodological consideration of our study. Our decision to use the provincial scale as the primary analytical unit was made deliberately, based on a comprehensive assessment of multiple factors, including data reliability, time-series consistency, and scale matching with key parameters (i.e., carbon density). The specific reasons are detailed below:

1. Data Reliability and Uncertainty:

While the spatially explicit (i.e., gridded) long-term Land Use/Cover Change (LUCC) datasets you mentioned do offer a high-resolution perspective, they are typically generated through techniques such as spatial downscaling or data fusion. This process inevitably introduces uncertainties stemming from model assumptions.

The reliability of such datasets is particularly challenging for historical periods. In contrast, provincial-level statistics are aggregated from long-term, relatively standardized administrative reporting systems. Although they have a lower spatial resolution, they represent the fundamental unit for historical land-use records in China and possess a high degree of reliability.

2. Time-Series Consistency and Continuity:

This study aims to construct a long-term inventory of carbon emissions from land use, for which data continuity and consistency are paramount. The data for the later period of our study were linked and calibrated with the "Second National Land Survey (2009)" and the "Third National Land Survey (2019)". Currently, the authoritative and fully open-access versions of these two surveys, which share a consistent statistical scope, are primarily available at the provincial summary level. Adopting the provincial scale thus maximizes the consistency of data sources and standards throughout the entire study period.

3. Scale Matching with Key Parameters (Carbon Density):

The core of our research is to estimate carbon emissions driven by land-use change, which requires coupling land-use area data with corresponding carbon density data. The carbon density datasets we employed—including data from the Second National Soil Survey of China (1979–1985), the China land ecosystem carbon density dataset by Xu et al. (2019), and the more recent Chinese Soil Series (since 2008)—are all fundamentally derived from field surveys and measurements at sample points. These sample points are spatially discrete and do not provide complete grid coverage. Therefore, aggregating both the land-use data and the carbon density sample data to the provincial scale is a more methodologically robust approach that ensures better compatibility between them.

In summary, while spatially explicit data offers advantages in displaying spatial patterns, we chose the provincial scale as it is the most appropriate and robust strategy for our research objectives. These objectives prioritize the construction of a long-term time series, the assurance of data reliability and consistency, and the scientifically sound coupling of land-use data with carbon density data derived from sample points.

This decision was a trade-off made after carefully evaluating the strengths and weaknesses of different data sources to ensure the accuracy and reliability of our final estimates.

Point 9. -Line 263. Provide references for "SAGE" and "PJ" datasets at first mention.

Response: Thank you for this comment. Revised.

For details, please see lines 192-193 of the revised manuscript.

Point 10. -Section 2.3.2 (Reliability Assessment). This section describes data sources rather than evaluating reliability. Move it to Section 2.2.1 (Land-use Data) for cohesion. If you put it in Methods, then the reliability results belong in Results/Discussion.

Response: We thank the reviewer for this insightful comment. Revised.

We agree that the content previously in Section 2.3.2 was a description of data sources and was misplaced. Accordingly, we have relocated this text to Section 2.2 Data sources to improve the manuscript's structure. This change is reflected in lines 147-210 of the revised version.

Point 11. -Line 500. Clarify key improvements in the latest bookkeeping model, including the updated disturbance-response curves, refined land-use transition rules? Or anything else?

Response: Thank you for this comment.

First, the new model refines the simulation of wood harvest to better reflect actual harvesting practices. The adjustments include correcting the post-harvest carbon allocation between 'wood products' and 'slash' to align with FAOSTAT data, and reducing the harvest intensity in secondary forests. This latter change necessitates simulating a larger harvested area to meet the same wood volume, thereby increasing the gross carbon sink from forest recovery.

Second, a key update is the proposal and simulation of alternative interpretations for 'Forest Conversion to Other Land' (FCO), a phenomenon observed in many tropical countries where the net loss of forest area exceeds the net gain in agricultural land. In contrast to the Houghton and Nassikas (2017) study which assumed a single pathway ('recovering forest'), this new research explores additional land-use conversion rules, including statistical error, 'shifting cultivation,' and 'degraded land,' to assess their distinct impacts on carbon emissions. This constitutes an in-depth refinement of the model's land conversion module and a robust uncertainty analysis.

Finally, this study did not update the response curves themselves (e.g., the rates and shapes of forest growth and soil carbon decomposition). The model continues to use prescribed, time-invariant response curves to simulate changes in per-hectare carbon stocks across different ecosystems and land-use change types.

In our research on long-term land-use carbon budget estimation for the China region, we adapted this model's framework to specifically address the issue of 'Forest Conversion to Other Land' (FCO), a topic detailed in our methodology (e.g., Sect. 2.3.2 Calculating annual land-use change). To resolve this, we analyzed the specific circumstances of FCO in China and selected combinations of response curves better suited to local characteristics. Consequently, while the fundamental forms of the response curves remain unchanged, their application rules for the China region were more thoroughly explored and refined.

Point 12. -Section 4.3 (Uncertainty Analysis) & Figure 11. Since you mention "2.3.4 Uncertainty assessment" in Methods, the Monte Carlo simulation results should appear in the Results. Reserve methodological limitations for the Discussion.

Response: Thank you for this comment. Revised.

We fully agree with your assessment. This is an excellent suggestion that significantly improves the logical structure of our manuscript and aligns it more closely with standard scientific writing conventions. Following your advice, we have restructured the paper by creating a new subsection, Section 3.3 'Uncertainty and Sensitivity Analysis,' within the results. We have moved the paragraphs detailing the results of the sensitivity analysis and Monte Carlo simulation, along with the corresponding Figure 11, from the Discussion to this new section. The remaining content discussing methodological limitations has been retained in the Discussion under the revised, more precise heading Section 4.3 'Limitations.' We are confident that these changes create a clearer distinction between our findings and their limitations (please see the revised manuscript, lines 480-516).

Point 13. -Data Implications for ESSD. As ESSD emphasizes data utility, expand on Applications: How can this dataset advance regional carbon budget assessments, climate modeling, or policy evaluations?

Response: Thank you for this comment. Revised.

Following your valuable suggestion, we fully acknowledge the importance of providing a more detailed discussion on the practical applications of our dataset. To this end, we have introduced a new section into the manuscript, Section 4.3 ("Implications and Applications"), to specifically elucidate how our dataset can facilitate future scientific research in the three critical areas of regional carbon budget assessment, climate modeling, and policy evaluation.Within this new section (lines 633-642), we have specifically detailed the following aspects:

1) For regional carbon budget assessment, the dataset provides a robust historical baseline for carbon fluxes from land-use change, enabling the separation of legacy emissions from contemporary fluxes. This is crucial for accurately attributing the drivers of the current terrestrial carbon sink and evaluating the effectiveness of ecological restoration efforts.

2) In climate and Earth system modeling, the dataset serves as an independent benchmark for evaluating and refining Dynamic Global Vegetation Models (DGVMs). Validation against the provincially-resolved emission estimates from this study can help constrain model parameters related to ecosystem responses to land-use change.

3) For policy evaluation, the dataset offers long-term quantitative evidence to assess the efficacy of land-use policies. The key transition from a carbon source to a sink around the 1980s strongly coincides with the implementation of China's large-scale ecological restoration policies, thus supporting the assessment of the potential effectiveness of such national-level interventions.

We are confident that this comprehensive elaboration has thoroughly addressed your concerns, effectively showcasing the scientific importance and practical value of our dataset for promoting frontier research in related fields.

Point 14. -Line 507. Table 2 should be Table 3?

Response: Thank you for this comment. Revised.

Point 15. -Regarding Figure 4: Is the term 'forest-grassland boundary' (林草界线) conventionally accepted? Suggest revising it to simply 'Eastern/Western China' for clarity.

Response: Thank you for this comment. Revised.