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 meticulously revised the manuscript. The main revisions include: 1) Based on your insightful comment, we have highlighted the methodological contribution of our carbon density analysis. 2) We have clarified the reason why the data table only includes land types that are directly relevant to our study's methodology. 3) We have updated the comparison data in the manuscript to reflect the latest research, including the Global Carbon Budget (GCB 2024), re-verified and explained the consistency between our data and the timing of the carbon sink transition, and, per your suggestion, added a discussion on the conceptual differences between the models.

We believe these revisions have further enhanced the rigor and clarity of the manuscript. Below are our point-by-point responses to your comments.

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.

We fully agree with your view that carbon density should be framed as a methodological contribution rather than a novel finding. 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.

Please allow us to provide a brief clarification. 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. 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 557–559, 562-565).

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 and the total value of +2.25 Pg C for 1980–2019 in Table 3 are 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 (lines 522–525).

We sincerely thank you for your two rounds of careful, professional, and timely reviews, which have led to a substantial improvement in the quality of our manuscript and a significant enhancement of its scientific rigor. Should you have any questions regarding our responses above, or additional comments and suggestions on the manuscript, please feel free to raise them during the third round of review, and we will make further careful revisions and improvements accordingly. Once again, we truly appreciate your dedicated efforts.

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 the second round of review. We concur that methodological validation, figure clarity, and systematic comparison are vital to our study. In response to your primary concerns regarding the construction of the land-use transition matrix, Figure 10 (Labeled as Figure 11 in the current version), and Table 3, we have undertaken substantial revisions to strengthen the study's scientific rigor.

In detail, our revisions include: (1) performing a comprehensive robustness check by implementing an alternative "area-weighted" allocation method, which validates the reliability of our land-use reconstruction; (2) updating the data sources and caption for Figure 10 (Labeled as Figure 11 in the current version) and supplementing the appendix with detailed definitions and origins of the datasets used; (3) clarifying the distinct inclusion criteria for Figure 10 (Labeled as Figure 11 in the current version) and Table 3, which serve different comparative aims, and adding an in-text explanation to prevent any potential confusion; and (4) refining the positioning of our results relative to existing literature based on newly included data. All changes are marked in the revised manuscript using the "track changes" feature, and specific line numbers are referenced in our detailed responses.

We are confident that these systematic improvements have thoroughly addressed the issues you raised, thereby enhancing the methodological soundness and credibility of our findings. Your expert guidance has been invaluable in improving the quality of our work, and we thank you once again.

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 this comment.

We thank the reviewer for their valuable and insightful comments, which have been instrumental in improving our manuscript. Your specific concern regarding the land use transfer allocation rule and its effect on the reliability of our findings is a critical point, which we have sought to address thoroughly.

Following your recommendation, we performed a comprehensive robustness check. We adopted an area-weighted allocation method—an objective, unbiased alternative—to recalculate the land use transfer matrices for the entire 1000–2019 study period. The rationale for this test has been added to the "Methods" section (revised manuscript, 2.3.3 Uncertainty assessment, lines 361–367).

Our systematic comparison of the original “priority-based” results with the new “area-weighted” results reveals high consistency. First, the absolute differences between the two sets of annual land use transfer matrices are minimal, as illustrated in a new heatmap (Appendix Fig C2), confirming the stability of our land use change reconstruction. Second, when the new land use data were used to re-estimate carbon emissions, the resulting trends and turning points were nearly identical to our original findings, with all new values falling completely within our previously reported uncertainty bounds (Figure 11).

Appendix C

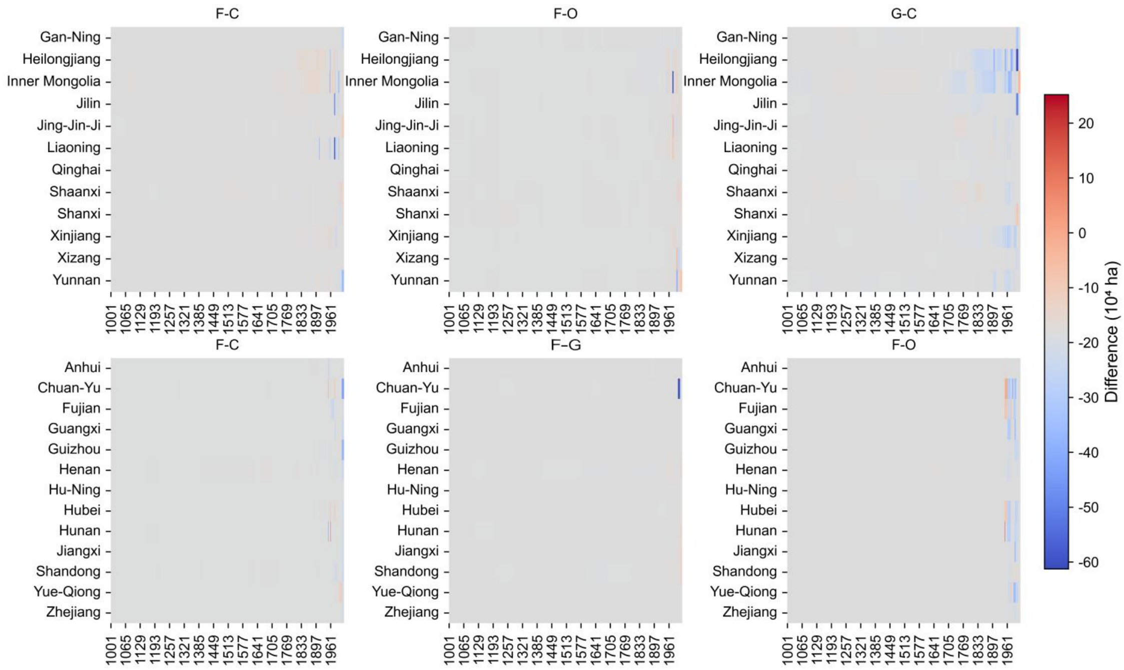

Figure C2 Differences in annual land-use transitions between the priority-based and area-weighted allocation methods. F-C denotes the conversion between Forest and Cropland, F-G represents Forest-Grassland conversion, F-O represents Forest-Other land conversion, and G-C represents Grassland-Cropland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

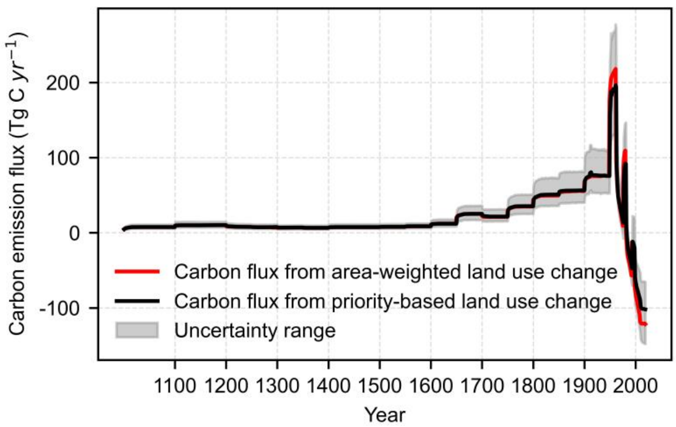

Figure 11. Comparison of carbon fluxes from land-use change using different calculation methods, with uncertainty assessment.

To integrate these results, we have added a new paragraph to the "Result" section

(3.3 Uncertainty and Sensitivity Analysis, lines 467–483), supported by the new figures. This analysis demonstrates that the core conclusions of our study are robust against the choice of allocation methodology. We are confident that this additional analysis directly addresses your primary concern and provides a more solid foundation for our conclusions.

Thank you again for your time and expertise. Your feedback has significantly improved the rigor of our study. We await your further review.

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 this comment.

1) We have updated the data from the Global Carbon Budget (GCB) 2019 to the latest version, GCB 2024.

2) We have amended the caption for Figure 10 (Labeled as Figure 11 in the current version) to clarify that the GCB estimate is not an independent data point. The caption now states that "The GCB estimate synthesizes the findings of Gasser (2020), Hansis et al. (2015), and Houghton (2023)." (lines 580–581)

3) The NGHGI.DB, NGHGI.DB.corrected, and TRENDYv8 datasets used in our study are all adopted from Obermeier et al. (2024) (*Obermeier, W. A., Schwingshackl, C., Bastos, A., Conchedda, G., Gasser, T., Grassi, G., Houghton, R. A., Tubiello, F. N., Sitch, S., and Pongratz, J.: Country-level estimates of gross and net carbon fluxes from land use, land-use change and forestry, Earth System Science Data, 16, 605-645, 10.5194/essd-16-605-2024, 2024.*). Detailed descriptions of these datasets, including their specific definitions, distinctions, and the correction procedures applied, are provided in Appendix Table D1. Specifically, the

TRENDYv8 dataset allows for the isolation of direct LULUCF impacts through the comparison of different scenarios (e.g., with and without land-use change). (Lines 551-555)

Appendix D

Table D1. Definitions and methodologies for the NGHGI.DB, NGHGI.DB.corrected, and TRENDYv8 datasets.

Dataset	Source and Description	Core Processing and Application
NGHGI.DB	National Greenhouse Gas Inventory (NGHGI) data reported by countries to the UNFCCC, with gap-filling applied.	Serves as the baseline data representing officially reported carbon fluxes from managed land.
NGHGI.DB.corrected	A corrected version of NGHGI.DB, adjusted to align with model-estimated anthropogenic fluxes.	Carbon fluxes from natural and indirect effects (e.g., CO_2 fertilization, climate change) are subtracted. This component is estimated by TRENDYv11 models under a scenario without land-use change.
TRENDYv8	Ensemble mean of nine Dynamic Global Vegetation Models (DGVMs).	Used to isolate the direct impacts of LULUCF by comparing results from different scenarios (e.g., with and without land-use change).

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 your valuable feedback. Your observation regarding the discrepancy between Table 3 and Figure 10 is very insightful, and you have correctly identified an important point that requires clarification. The two formats were intentionally designed for different comparative purposes, which we are happy to explain here. We will also add a note to the manuscript to clarify this for readers.

Table 3 and Figure 10 are not merely different presentations of the same data; they are complementary comparisons targeting two categories of studies with distinct data attributes. Our criteria for inclusion were as follows:

- Figure 10 Inclusion Criterion: This figure compiles results from all studies for which we could obtain annual-resolution time-series data. This format allows for a direct, year-by-year visual comparison of dynamic trends. Consequently, it includes multiple Dynamic Global Vegetation Models (DGVMs), bookkeeping models driven by remote sensing or global datasets (e.g., HYDE/LUH), and annual data from NGHGI/FAOSTAT.
- Table 3 Inclusion Criterion: This table focuses on key studies that did not provide publicly available annual time-series data but reported total or average estimates over specific periods. These studies, particularly the foundational works by Yang et al. (2023), Yang et al. (2019), Li et al. (2014), and Ge et al. (2008), are crucial for understanding the historical carbon budget of China based on century-scale land-use reconstructions. We present their core findings in the table to facilitate a direct comparison of their period-aggregated results.

Based on these principles, we also included two specific studies in Table 3 for the following reasons:

- Houghton and Castanho (2023): This study was included because we used a bookkeeping model identical to theirs, whereas several other studies in the table used earlier versions of this model. Including it provides a direct methodological benchmark for comparison.
- Yu et al. (2022): Although this is a DGVM study with time-series data (as shown in Fig. 10), we also included it in Table 3 because it shares a critical attribute with the other studies in the table: it uses a century-scale, historically reconstructed land-use dataset specifically for the China region as input. This distinguishes it from most studies in Fig. 10 that rely on global datasets, making its input data and resulting estimates highly comparable to the other works listed in Table 3.

We hope this explanation clarifies the design rationale for Figure 10 and Table 3. To prevent any confusion for future readers, we will add a concise explanation to the manuscript (Lines 525-529). Thank you again for your insightful comments and suggestions.

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 insightful feedback. You are correct, and we agree that in the previous version of Figure 10 (Labeled as Figure 11 in the current version), our post-2000 estimates were among the most negative values (the lowest, excluding NGHGI and FAOSTAT data), not at an "intermediate level." Your observation was spot-on.

In this revision, per the suggestion of Reviewer 1, we have incorporated the recent study by Zhu et al. (2025) (Zhu, Y., Xia, X., Canadell J., Piao, S., Lu, X., Mishra, U., Wang, X., Yuan, W., and Qin, Z.: China's Carbon Sinks from Land-Use Change Underestimated. Nature Climate Change, 15, 4: 428-35, 2025.) into Figure 10 (Labeled as Figure 11 in the current version). The findings from this study indicate an even larger carbon sink between 1992 and 2020. Consequently, when compared against this new dataset, our estimates are no longer the lowest in the range.

We have revised the relevant text in the manuscript to accurately reflect this updated comparison. Please see Lines 557-559 for the specific changes.

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. We have carefully considered the issue of inconsistent statistical calibers in our data integration and would like to provide a detailed explanation here.

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 608-619 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)

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. 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 640-649).

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. 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.

Thank you again for your guidance.

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.

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.

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. 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.

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 467-492).

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.

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 597-607), 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.