

Response to Reviewer 1

Our comments are inset in blue colour following each point of the reviewers. The text quoted directly from the revised manuscript is set in italics. The line numbers cited in our response refer to the revised manuscript with no changes marked.

The authors create a high resolution bottom-up inventory of methane emissions in China, offering a valuable dataset for the community. My comments are concerned primarily with (1) clarifying methodology and (2) strengthening evaluation against other datasets. I support publication after these comments are addressed.

Response: We sincerely appreciate the Reviewer's valuable comments and suggestions, which help us significantly improve the manuscript. In the revised manuscript, we addressed each of these concerns: 1) enhancing the methodological descriptions with particular attention to handling missing data; 2) conducting a comprehensive comparisons of CHN-CH₄ with EDGAR (v8), PKU-CH₄ (v2) and GFEI spatially; 3) performing an extended evaluation using 26 bottom-up estimates and 14 top-down atmospheric inversions to validate CHN-CH₄ at both sectoral and national levels for 1990-2020; 4) thoroughly discussing uncertainty sources while refining all figures and clarifying text throughout.

Major comments:

1. Figure 3: It is not clear to me what is being compared in the scatter plots d-i. Are these all individual grid cells in China for a given year (2000/2019)? For 2000-2009, is this a decadal average or all data for each year? It is also clear that there is a density plot, but there is no colorbar showing how many points are combined at the yellow saturation. Perhaps this would be better shown on a log-log plot so lower emissions grid cells are more visible and easier to compare.

Response: In the initial submission, the data for 2000–2009 represents annual average across all data, and we displayed only a randomly selected subset of 10,000 points for clarity because of the low emission values in many grid cells. As suggested, we improved the figure and adopted a log-log plot to better illustrate the differences between our results and those from EDGAR, PKU-CH₄, and GFEI (please see Figure 3 and Figure 4 in the revised manuscript).

2. Lines 307-311: How do you know EDGAR is overestimating these emissions? Bottom-up

estimation is complicated and can have errors, and top-down quantification similarly has errors; our uncertainties are generally quite large. In particular, I disagree with line 311; I think this paper would greatly benefit from a more detailed comparison with EDGAR at sector, regional, and national levels. Disagreement between your estimates is expected, but it is important to understand why your inventory disagrees.

Response: We acknowledge that bottom-up estimates inherently carry high uncertainties, and thus it would be oversimplistic to categorically assert an overestimation in EDGAR. In the revised manuscript, we provided a more comprehensive comparison between CHN-CH₄ and EDGAR across pixel, sectoral, regional, and national levels. Spatially, CHN-CH₄ exhibits strong agreement with EDGAR, with most grid cells aligning closely along the 1:1 line (Figure 3 below). A detailed histogram analysis further reveals that spatial differences are predominantly centred around zero. However, notable deviations still exist. CHN-CH₄ exhibits higher emissions in North China (e.g., Shandong and Henan) but lower estimates in energy-intensive provinces (e.g., Shanxi and Sichuan), major rice-growing regions (e.g., Hunan and Jiangxi), and developed coastal areas. At the sectoral and national levels, we recollected 40 existing inventories (26 bottom-up estimates and 14 top-down inversions) for comparison. CHN-CH₄ provides continuous coverage from 1990 to 2020 at a higher spatial resolution than most references. In terms of total anthropogenic emissions, EDGAR exceeds CHN-CH₄ by over 36% and PKU-CH₄ by over 30%, primarily due to overestimations in rice cultivation and wastewater sectors. For other sectors, EDGAR's magnitude and variability align well with the reference inventories. Given these findings, we refrained from conclusively attributing overestimation to EDGAR and revised our manuscript accordingly. The updated text now reads (Lines 309-315):

Spatial validation against EDGAR v8 and PKU-CH₄ v2 datasets for the years 2000, 2009, and 2019 demonstrates strong agreement with CHN-CH₄ along the 1:1 line (Fig. 3). The higher R² between CHN-CH₄ and EDGAR indicate greater spatial consistency compared to PKU-CH₄. Notable regional differences persist, with CHN-CH₄ showing lower emission estimates than both EDGAR and PKU-CH₄ in energy-intensive provinces (e.g., Shanxi and Sichuan) and major rice-growing regions (e.g., Hunan and Jiangxi). Particularly, the deviations from EDGAR are more pronounced than those from PKU-CH₄.

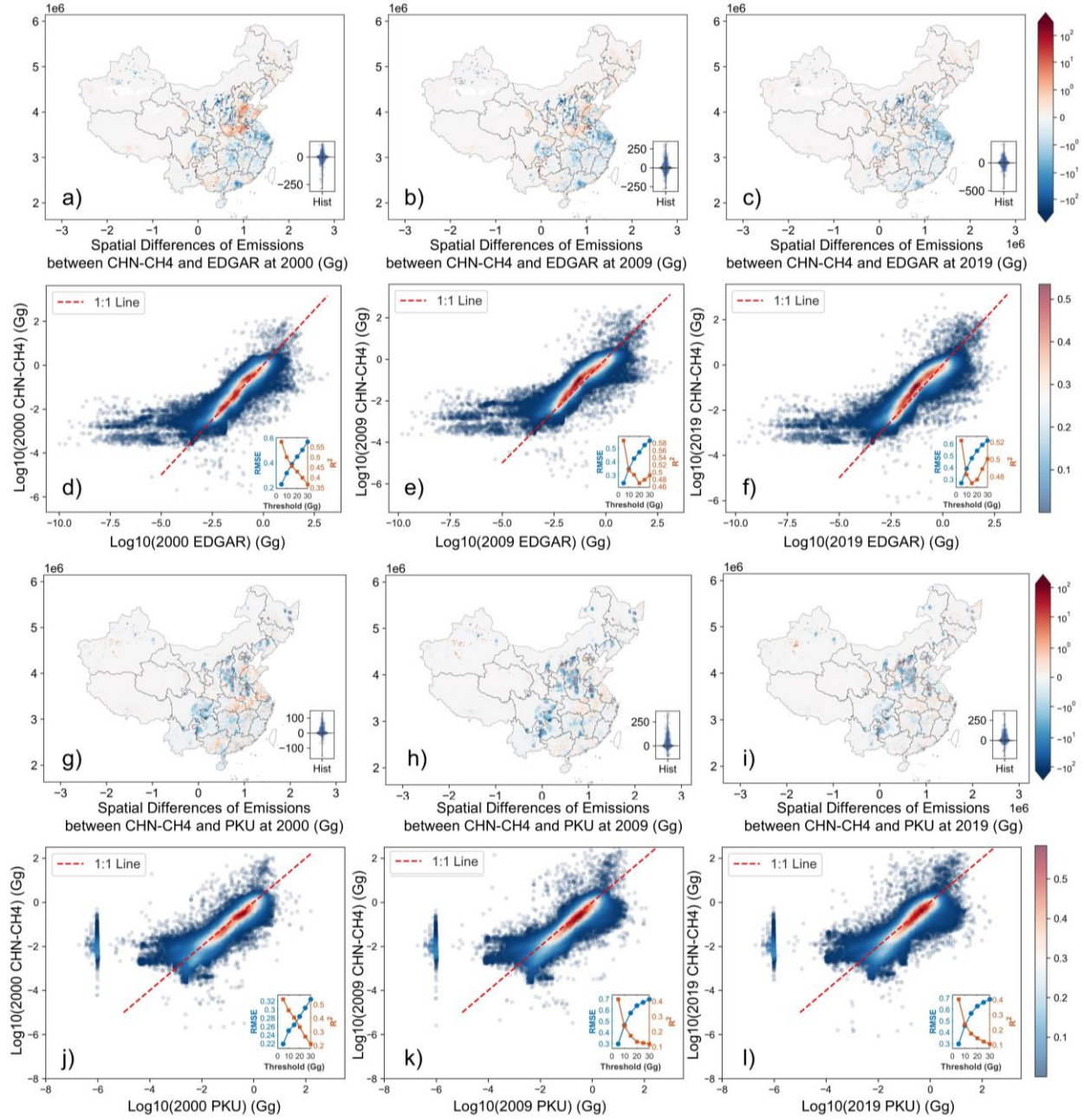


Figure 3. Pixel-level comparisons between CHN-CH₄ and EDGAR/PKU-CH₄. a-f) represent the comparison between CHN-CH₄ and EDGAR v8, while g-l) represent the comparison between CHN-CH₄ and PKU-CH₄ v2 at year 2000, 2009, and 2019. ‘Hist’ in each spatial map represents the histogram of the differences between CHN-CH₄ and EDGAR/PKU-CH₄, with the unit Gg. The bottom-right subfigure in each log-log plot presents threshold-dependent performance metrics, demonstrating how RMSE and R^2 vary when excluding grid cells over specific emission thresholds.

Lines 338-349 read as:

26 bottom-up inventory estimates and 14 top-down atmospheric inversion models were selected for the sectoral and national comparisons (Fig. 5). The CHN-CH₄ dataset reveals a clear increasing trend in total anthropogenic methane emissions from 1990 to 2020, though with moderate agreement to reference data ($R^2 = 0.11$, RMSE = 10.5 Tg, MAE = 7.6 Tg) (Fig. 5a). This discrepancy is largely attributable to EDGAR's systematic overestimation, which exceeds CHN-CH₄ estimates by at least 36% throughout the study period - consistent with the 30-40% overestimation previously reported by Peng et al. (2016). When EDGAR is excluded, the upscaled national emissions show substantially better agreement ($R^2 = 0.57$, RMSE = 6.7 Tg, MAE = 4.7 Tg). The overestimation is particularly pronounced in the rice cultivation and wastewater sector, potentially due to EDGAR's use of higher emission factors and assumed continuous flooding rates. In other sectors, this phenomenon is not obvious. Notably, most inventories (including CHN-CH₄) show a gradual decline post-2014, although EDGAR's estimates continued increasing with at a slower rate.

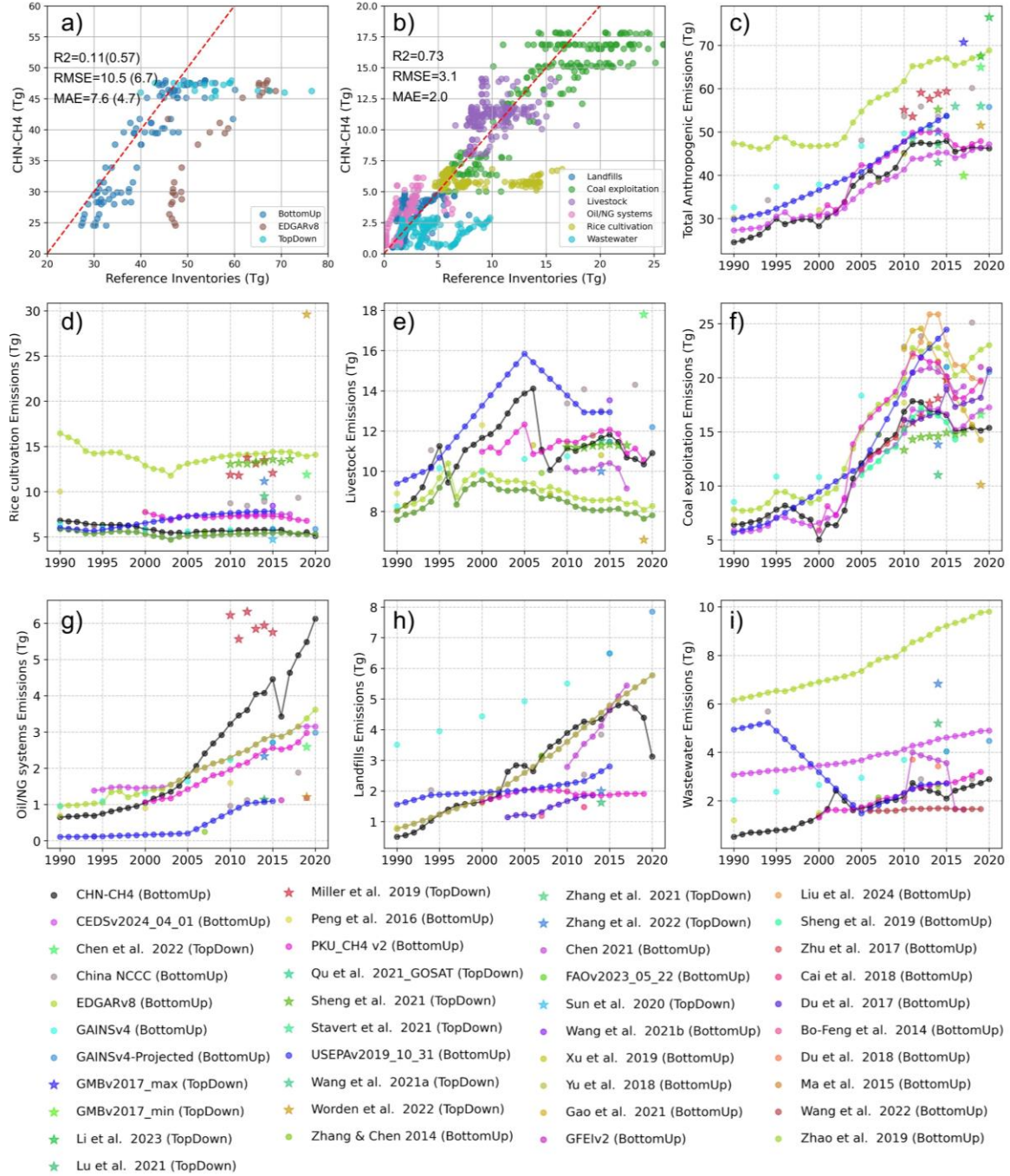


Figure 5. Sectoral and national comparisons between CHN-CH4 and reference inventories. a) National-level emission comparisons between CHN-CH4 and references, b) Combined sectoral emissions comparison across all inventory sources, c) Variations of total anthropogenic emissions of CHN-CH4 and references, and d-i) Individual sector-specific variations between CHN-CH4 and the reference (rice cultivation, livestock, coal exploitation, Oil/NG systems,

landfills, and wastewater, respectively). The red line is 1:1 line.

3. In addition to your scatterplots in Figure 3, I'd also like to see a spatial comparison with other inventories, especially EDGAR and also perhaps some leading sectoral inventories like GFEI for fossil fuels (especially coal). It would also be interesting to see spatial comparisons with a top-down posterior estimates. Do your maps generally agree on the spatial patterns of emissions, or are there significant differences? This could perhaps be shown in a difference plot like Figure 5.

Response: In the revised manuscript, we enhanced the spatial analysis with three bottom-up inventories (GFEI, EDGAR, and PKU-CH₄). Grid-cell-level comparisons with top-down posterior estimates were not included due to data accessibility constraints. Instead, we conducted comprehensive sectoral and national-level comparisons using published top-down estimates from the literature (see Figure 5). This multi-scale evaluation strengthens the robustness of our CHN-CH₄ emission estimates. For the three bottom-up inventories, we made comparison across three energy sectors: (1) total fuel exploitation, (2) coal exploitation, and (3) oil/natural gas (NG) exploitation (see Figure 4 below). Results demonstrate well spatial consistency between CHN-CH₄ and GFEI for total fuel exploitation (with differences clustered around zero), while the main discrepancies originate from coal mining (RMSE: 39 Gg; MAE: 24.5 Gg). In contrast, the oil/NG sector demonstrates much closer alignment, as evidenced by the log-log plot. Among the three inventories, EDGAR exhibits the best spatial distribution agreement with CHN-CH₄ (with data points densely concentrated along the 1:1 line), whereas PKU-CH₄ provides closer emission magnitude estimates (lower RMSE and MAE values). We added these descriptions in the revised manuscript. The updated text now reads (Lines 324-334):

Further comparison with the Global Fuel Exploitation Inventory (GFEI) across energy sectors shows well spatial agreement for oil/NG systems but larger discrepancies in coal mining (RMSE > 39.4 Gg, MAE >24.5 Gg) (Fig. 4). In contrast, the oil/NG sector demonstrates much closer alignment, as evidenced by the log-log plot (Fig. 4c and 4f). Compared to GFEI and PKU-CH₄, CHN-CH₄ exhibits stronger spatial consistency with EDGAR, as evidenced by a higher concentration of points along the 1:1 line. The spatial discrepancies between GFEI and EDGAR primarily arise from differences in their underlying data sources for fuel exploitation; EDGAR's coal and oil/gas system locations were adopted in this study. Despite these variations, the overall alignment supports the reliability of CHN-CH₄ in identifying key sectoral and regional emission

patterns.

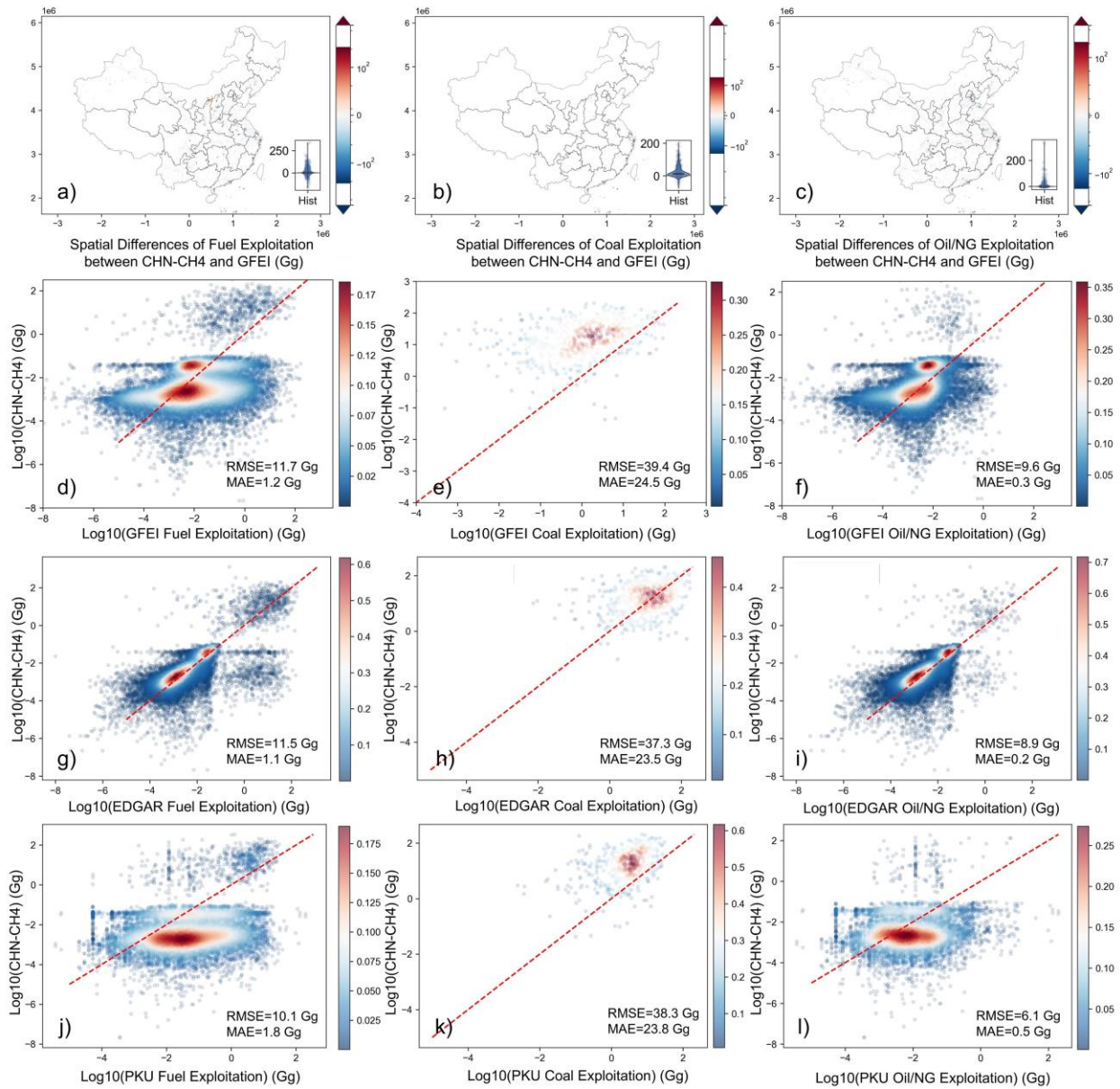


Figure 4. Pixel-level comparisons between CHN-CH4 and three inventories (GFEI, EDGAR, and PKU-CH4) in the energy sectors for 2019. a-c) show spatial differences between CHN-CH4 and GFEI. d-i) present log-log scatterplots of pixel-level emissions: CHN-CH4 versus GFEI (d-f), EDGAR (g-i), and PKU-CH4 (j-l), respectively.

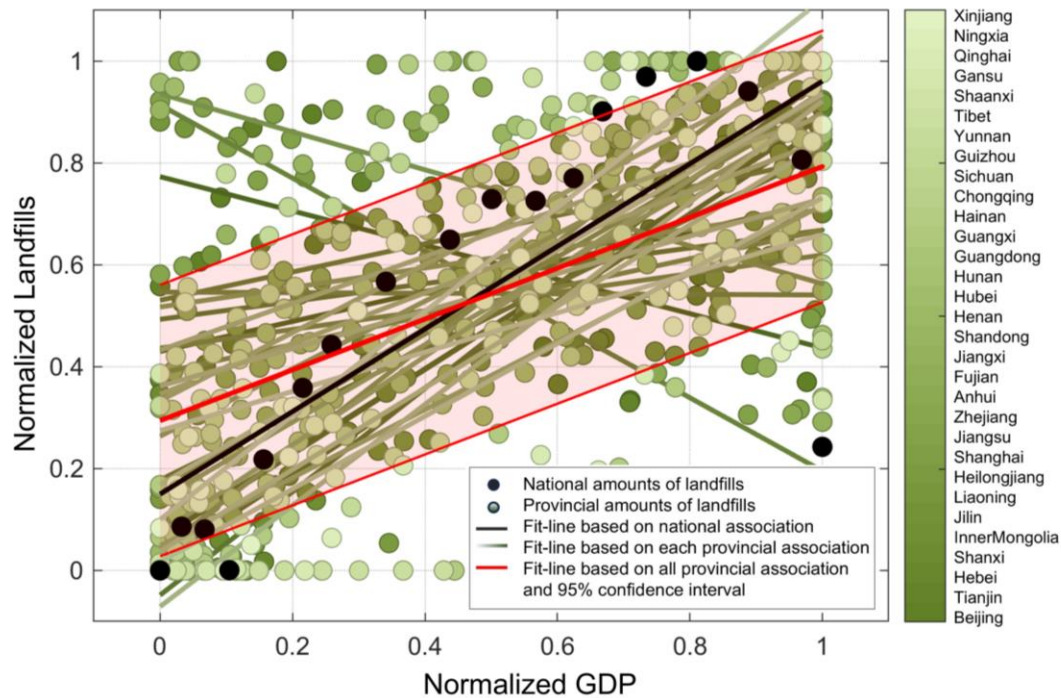
4. On that note, what version of EDGAR do you use for your comparison? Table 1 suggests V8 but it isn't clear from the citations. How do your comparisons look with different versions of EDGAR? V8 for example tends to have more emissions from point sources.

Response: We compared the latest version of EDGAR (v8), which incorporates more detailed point-source emissions, particularly for industrial processes related to fossil fuel use. Due to limited spatial activity data for industrial emissions in China, we allocated the fossil fuel consumption data using economic proxies such as GDP and population density, rather than distinguishing specific industrial processes (e.g., chemical or metal production). For spatial comparison with CHN-CH₄, we followed the IPCC 2006 guidelines and aggregated EDGAR's anthropogenic emissions to ensure consistency in categorization, including agriculture sectors (*AGS - Agricultural soils, AWB - Agricultural waste burning, ENF - Enteric fermentation, and MNM - Manure management*), energy sectors (*PRO_COAL - Fuel exploitation COAL, PRO_GAS - Fuel exploitation GAS, PRO_OIL - Fuel exploitation OIL, REF_TRF - Oil refineries and Transformation industry, TRO - Road transportation, TNR_Other - Railways, pipelines, off-road transport*), and waste sectors (*WWT - Waste water handling, and SWD_LDF - Solid waste landfills*). Since our analysis did not include emissions from solid waste incineration, this sector was excluded from the waste-related comparisons. In the revised manuscript, we have expanded the details of our comparison with EDGAR to enhance clarity and minimize potential ambiguities.

5. I would like more information and discussion of the extrapolation and interpolation methodologies used in this paper. It seems like the authors are dependent on simple linear regression for e.g. landfills (255-261), regressing against GDP to predict landfills before yearbook data begin in 2003. I see that on a national basis the correlation in Figure S2 is strong, but is this also true at the provincial level? Similarly, could the authors clarify how missing data are handled, across all sectors? Are the methods the authors use to handle these data in line with previous work?

Response: We have expanded the methodology and discussion sections to provide better clarity regarding our approach to landfills sector. Since the gridded activity data is unavailable, we adopted the methodology of Peng et al. (2016) to allocate provincial statistical data to gridded cells using linear relationships. However, since provincial landfill data were only available from 2003 onward, we employed the following approach to address missing data for earlier years (1990-2002). First, we estimated national landfill amounts by establishing the linear relationship with GDP. As shown in the figure below, the national trend (black line) falls within the 95% confidence interval of provincial relationships (red line), demonstrating that the national pattern effectively captures provincial variations. In this case, we applied the national GDP-landfill relationship to estimate

1990-2002 national totals, and then allocated these estimates to provinces using their 2003-2008 average contribution to national landfill amounts. Later, the provincial landfills for the whole period (1990-2020) were distributed to grid cells using the spatial distribution of GDP, and then the corresponding methane emissions were calculated from the gridded data.



The methods to handle missing data in landfills were in line with previous work (see Figure 2 in the manuscript). Our analysis began by establishing temporal anchors at the earliest (t_1) and most recent years (t_2) with available statistical data for each province and sector. We employed linear interpolation to estimate missing values between these anchor years based on historical trends. For periods preceding t_1 and following t_2 , national-level statistics were prioritized, and then provincial data were allocated according to their average proportion to the national total over the nearest five-year period. This five-year period was chosen to align with China's Five-Year Plan cycles, given the relative stability of economic development policies. This approach provides a reasonable estimation method when detailed historical provincial data are unavailable, while maintaining alignment with national totals and provincial development patterns.

Minor comments:

6. Could you describe the methodology in the left portion of Figure 1 in more detail (at bottom

left)? I don't understand what t1 and t2 are, for example.

Response: t1 and t2 represents the earliest and most recent available year that we can get the data from statistical yearbook. In the revised manuscript, we added more information to describe the figure (please see Lines 79-83).

7. Abstract (and elsewhere, e.g. line 358 and Table 5): Do you really have 3 decimal points of confidence in your estimates? I recommend rounding.

Response: We agree. As most literature have done, we have adopted one decimal place for reporting confidence values throughout the manuscript. We have carefully reviewed the entire document to ensure consistency in this reporting format and to eliminate any potential ambiguities in our presentation of results.

8. From your dataset, I see that your data is at annual temporal resolution, but this is not clear in the text of the paper. Could you clarify this?

Response: We modified the sentence in the introduction section to clarify the temporal resolution of CHN-CH₄ (please see Lines 61-64).

9. Throughout, e.g. lines 182-183, 224, 310: Are the years 1980-2010 correct here or do you mean 1990-2020? If 1980-2010 is correct, how do you handle 2010-2020?

Response: We have corrected the time period from our initial version to the accurate range of 1990–2020 throughout the entire manuscript, including all relevant figures, tables, and methodological descriptions.

10. Line 234: EF for coal combustion is missing

Response: We added the EF for coal combustion, $1\text{ kgCH}_4\text{TJ}^{-1}$ as suggested by IPCC (2006).

11. Line 260-261: The authors grid landfill data using GDP. I am not familiar with landfills in China but I wonder if this a reasonable assumption? Perhaps the authors could compare with estimates that incorporate more spatial information, like this one: Bofeng Cai et al. CH₄ mitigation potentials from China landfills and related environmental co-benefits. Adv.4,eaar8400(2018).DOI:10.1126/sciadv.aar8400

Response: While we acknowledge the inherent uncertainties, our assumption of using GDP as a proxy for landfill CH₄ emissions appears reasonable given the well observed correlation between

national/provincial landfill data and GDP, particularly in the absence of precise geolocated landfill data. The Reviewer's reference to prior literature (reporting 1.84 Mt CH₄ from landfills in 2012 from Chinese Ministry of Environmental Protection, provincial environmental protection bureaus and field investigations) highlights an important discrepancy with our estimate of 4.2 [3.7-5.1] Mt. This substantial difference suggests that conventional bottom-up approaches may systematically misestimate emissions compared to observational constraints. We have expanded our discussion of this finding in the revised manuscript. The updated text now reads (Lines 464-471):

Instead of using specific locational data, the distribution of GDP and population is employed for regression and prediction purposes, which may not accurately reflect the true spatial distribution of emissions. This reliance on indirect proxies can introduce significant uncertainties in the spatial allocation of emissions. For instance, our estimate of methane emissions from landfills in China for 2012 (4.2 Mt) differs from the 1.8 Mt reported by Cai et al. (2018), who estimated the emissions using the activity data from Chinese Ministry of Environmental Protection, provincial environmental protection bureaus and field investigations. This substantial difference suggests that conventional bottom-up approaches can systematically misestimate emissions compared to observational constraints.

12. Figure captions in general should be more descriptive of figure content.

Response: We have carefully revised all figure captions in the manuscript to ensure they provide clearer descriptions and remove ambiguous language to better align with the visual presentation.