RC2

The paper compares six CO_2 emission inventories for China from 2000 to 2023, including global inventories (ODIAC, EDGAR, GEMS) and China-specific ones (MEIC, CHRED, CEADs). It highlights large differences between inventories, especially EDGAR vs. MEIC, and differences in spatial distributions. This is important because China has ambitious carbon reduction goals, so accurate quantification of CO_2 emissions is essential for policy and climate modelling. The paper fits within the journal's scope as it addresses atmospheric emissions and their uncertainties.
Limitations of this review: I am not an expert in CO_2 emissions inventories and the relevant literature, so my comments focus on interpretation, clarity, and presentation rather than technical accuracy of methods.

Major comments

The paper is well-structured, the argument is easy to follow, and the language is clear. However, the following aspects would need to be addressed before publication.

1. Clarify the novelty of the study

It is unclear how this work differs from previous studies. Is the novelty in using updated versions of inventories, applying new harmonisation methods, or drawing new conclusions? Please add a short paragraph in the introduction explicitly stating what is new compared to other studies mentioned (e.g., Han et al., 2020a; Liu et al., 2015; L. Zheng et al., 2025).

Response: We appreciate the reviewer's valuable comment. To highlight the novelty of our work, we have added a paragraph in the Introduction and a more detailed paragraph in the Conclusion explicitly outlining the main advancements compared with previous studies. Specifically, this study (1) extends the temporal coverage to 2000–2023 and identifies three distinct emission phases reflecting changes in energy policy and structure; (2) evaluates internal inconsistencies within CEADs and recommends using CEADs (sectors) for provincial analyses; (3) reveals significant sectoral spatial allocation differences, particularly between EDGAR and MEIC in the transport sector; (4) quantifies scale-dependent uncertainties, showing that provincial uncertainty (CV) is two to ten times higher than national uncertainty; and (5) demonstrates that CEADs and MEIC yield consistent estimates across nine representative provinces. At the national scale, CAMS exhibits the smallest deviation from the National Greenhouse Gas Inventory (NGHGI), while ODIAC aligns most closely with the six-inventory mean during the study period. These clarifications have been added to Section 1 to highlight the study's novelty and rationale for using the latest inventory versions, and to Section 4 to summarize the new insights contributions.

Revision:

(1) ***Section 1, paragraph 5:*** *"To this aim, this study conducts a comprehensive analysis of the spatiotemporal variation of China's anthropogenic CO_2 emissions and investigates the differences among six widely used emission inventories at their latest versions: the global inventories ODIAC, EDGAR, MEIC, GEMS, CAMS, and the China-specific inventory CEADs. The data and methods*

are presented in Section 2. We report our results in Section 3 and conclude the paper in Section 4. Compared with previous studies (Han et al., 2020b; Zheng et al., 2025), we extend the temporal coverage to 2000-2023, enabling a more current and consistent assessment of recent emission trends, inter-inventory discrepancies, and scale-dependent uncertainties across China. "

(2) ***Section 4, paragraph 5:*** *"In summary, this study extends previous work by identifying a three-phase trend in China's anthropogenic CO_2 emissions from 2000 to 2023 and quantifying the emission uncertainties (1σ) at both national and provincial levels. At the national level, CAMS shows the closest agreement with the government-reported NGHGI, while ODIAC aligns best with the multi-inventory mean over the study period. At the provincial level, the Chinese local inventories, CEADs and MEIC, provide the most consistent estimates for regional studies. Differences in spatial proxies significantly affect the spatial distribution of sectoral emissions, as shown by the contrasting transport emission patterns in EDGAR and MEIC. We also clarify the appropriate use of CEADs for provincial analyses. Our results further underscore the importance of improving the consistency of regional inventories to provide a stronger scientific basis for China's emission mitigation and carbon neutrality policies."*

2. Recommendations for users

The conclusion clearly summarises findings but could be strengthened by adding actionable guidance. Readers would benefit from answers to the following questions:

- Which inventories are most reliable for specific applications?
- What are the main uncertainties that remain?
- How can inventory producers improve the next inventory versions?

A summary table of findings and recommendations could make this section more impactful.

Response: We thank the reviewer for these constructive suggestions. Providing practical recommendations would strengthen the manuscript, and we have revised the text to improve clarity for readers.

(1) Consistency assessment at national and provincial levels

We agree that identifying which inventories are more reliable is crucial. However, determining the absolute accuracy of each inventory requires direct comparison with independent observations (e.g., atmospheric CO_2 measurements together with an inversion model), which is beyond the scope of this study. Therefore, in this study, we mainly assessed the internal consistency of the six inventories and their deviations from independent references. Specifically, we included the National Greenhouse Gas Inventory (NGHGI) reported by the Chinese government to the UNFCCC for national-level comparison. Our results show that CAMS exhibits the greatest consistency with the NGHGI, while ODIAC aligns most closely with the six-inventory mean. At the provincial level, uncertainties are 2-10 times higher than at the national scale. Although absolute references remain uncertain, CEADs and MEIC demonstrate strong agreement across nine representative provinces, particularly in Inner Mongolia, Shandong, Henan, Hubei, and Shanghai.

(2) Main source of uncertainties

Different downscale methods and spatial proxies might be the primary source of uncertainties across inventories. This is quantitatively supported by our finding that the uncertainties at provincial level are two to ten times higher than at the national level. Furthermore, our analysis shows that differences in spatial proxies significantly affect the spatial distribution of sectoral emissions, as shown by the contrasting transport emission patterns in EDGAR and MEIC.

(3) Recommendations to improve inventory reliability

To enhance the reliability of future inventory versions, we recommend enhanced cross-validation with national statistics and transparent documentation of proxy methodologies. In addition, expanding ground-based and satellite observations would enable comprehensive independent validation. CO_2 flux measurements can be directly compared with bottom-up estimates, while atmospheric CO_2 mole fractions measurements, when integrated with inversion model, yield top-down emission estimates. These top-down results can then be systematically compared with bottom-up inventories to identify discrepancies across regional and national scales.

Revision:

(1) ***Section 3.1, paragraph 2:*** *"To further assess the consistency of the six inventories, we calculate the mean absolute difference (MAD), which is defined as the multi-year mean of annual absolute differences between each inventory and either the NGHGI or the six-inventory mean. Compared with NGHGI, the MADs range from 0.156 Gt year^{-1} (CAMS) to 0.835 Gt year^{-1} (MEIC). Against the six-inventory mean, the MADs range from 0.12 Gt year^{-1} (ODIAC) to 0.449 Gt year^{-1} (MEIC). EDGAR reports the highest emissions, which is about 0.370 Gt year^{-1} larger than the mean emission. MEIC shows the lowest emission levels, which is about 0.449 Gt year^{-1} less than the mean emission. Overall, CAMS exhibits the greatest consistency with the NGHGI, being at least 30% lower than that of the other inventories. In comparison, ODIAC agrees most closely with the six-inventory mean, with an MAD at least 58% lower than the others."*

(2) ***Section 4, paragraph 1:*** *"China's annual anthropogenic CO_2 total emission increases from 3.42 Gt in 2000 to 12.03 Gt in 2023. When compared with the officially reported NGHGI and the six-inventory mean, CAMS shows the smallest deviation from the NGHGI, while ODIAC agrees most closely with the multi-inventory mean. The six inventories display a broadly consistent emission trend, but their discrepancies among the inventories have widened from 0.41 Gt year^{-1} to 1.63 Gt year^{-1}, …"*

(3) ***Section 4, paragraph 4:*** *"…The pronouncedly higher emissions in the coastal megacities (e.g., Shanghai, Jiangsu, and Guangdong) by ODIAC and the abnormal increase in CAMS by 50-230% in Liaoning, Hubei, and Shanghai exacerbate this divergence. Despite these inconsistencies, CEADs and MEIC exhibit broadly consistent estimates across nine provinces, especially in Inner Mongolia, Shandong, Henan, Hubei, and Shanghai."*

(4) ***Section 4, paragraph 5:*** *"In summary, this study extends previous work by identifying a three-phase trend in China's anthropogenic CO_2 emissions from 2000 to 2023 and quantifying the emission uncertainties (1σ) at both national and provincial levels. At the national level, CAMS shows the closest agreement with the government-reported NGHGI, while ODIAC aligns best with the multi-inventory mean over the study period. At the provincial level, the Chinese local inventories, CEADs and MEIC, provide the most consistent estimates for regional studies. Differences in spatial proxies significantly affect the spatial distribution of sectoral emissions, as shown by the contrasting transport emission patterns in EDGAR and MEIC. We also clarify the appropriate use of CEADs for provincial analyses. Our results further underscore the importance of improving the consistency of regional inventories to provide a stronger scientific basis for China's emission mitigation and carbon neutrality policies."*

(5) ***Section 4, paragraph 6:*** *"Overall, reliable emissions quantification requires scale-appropriate inventories (e.g., the sectoral CEADs emissions versus the province-based CEADs emissions), improved spatial proxies (e.g., CPED vs. CARMA), and ensemble approaches to mitigate biases, especially in the carbon-intensive eastern regions. It should be noted that this study lacks an observational benchmark to assess these inventories. Future efforts should incorporate direct flux measurements or top-down emissions derived from inversion modeling, in combination with CO_2 mole fraction observations, to compare and constrain bottom-up inventories at both regional and national scales."*

Section 3.1, Figure 1:

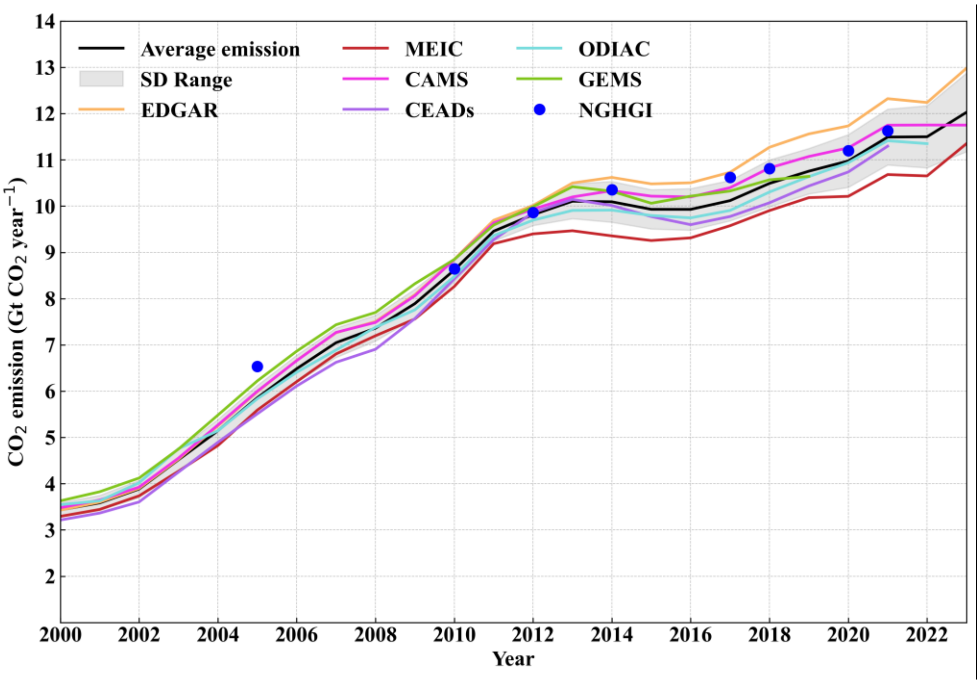


Figure 1. *Annual anthropogenic CO_2 emissions in mainland China from 2000 to 2023, as reported by six emission inventories: EDGAR, MEIC, CAMS, CEADs (up to 2021), ODIAC (up to 2022), and GEMS (up to 2019), and one government-reported data (NGHGI). Apart from ODIAC, all inventories provide national totals directly. We calculated China's emissions by summing the grid values within China for ODIAC. The shaded area indicates the standard deviation of the six inventories. It's noteworthy that the inter-inventory mean and SD were calculated from the above mentioned six inventories.*

3. Comparison to observations

The study compares inventories against each other. Without observational benchmarks, it is difficult to assess which inventory is closer to reality. Could you explain why observational comparisons were not included? If data limitations prevented this, could you state them explicitly and discuss implications for interpreting results?

Response: We appreciate the reviewer's insightful comment. We fully agree that observational benchmarks are essential for evaluating the accuracy of emission inventories. However, such comparisons were not included in this study due to data limitations. Direct CO_2 flux measurements, such as those from eddy covariance or mass balance, are spatially sparse and only represent local scales. Consequently, they are unsuitable for evaluating national or provincial emission totals. In addition, fluxes derived from atmospheric inversion model together with CO_2 mole fraction measurements can provide valuable top-down constraints but are also strongly affected by available data and model uncertainty. Therefore, incorporating these datasets would not provide consistent national-scale evaluation between 2000-2023.

Our study focuses on assessing the internal consistency among inventories and their deviations from independent references (i.e., the National Greenhouse Gas Inventory, NGHGI). We acknowledge that the absence of an observational benchmark limits the ability to identify which inventory is closer to reality. Future work should integrate direct flux observations and top-down emissions from inversion modeling to independently evaluate and constrain bottom-up inventories at both regional and national scales. We have added a clarification in Section 4, paragraph 6 to address this point.

Revision:

Section 4, paragraph 6: *"Overall, reliable emissions quantification requires scale-appropriate inventories (e.g., the sectoral CEADs emissions versus the province-based CEADs emissions), improved spatial proxies (e.g., CPED vs. CARMA), and ensemble approaches to mitigate biases, especially in the carbon-intensive eastern regions. <u>It should be noted that this study lacks an observational benchmark to assess these inventories. Future efforts should incorporate direct flux measurements or top-down emissions derived from inversion modeling, in combination with CO_2 mole fraction observations, to compare and constrain bottom-up inventories at both regional and national scales.</u>"*

Specific comments

1. Line 80: MEIC is described as China-specific but later implied to be global. Could you clarify?

Response: We thank the reviewer for pointing out this potential confusion regarding the MEIC inventory. We acknowledge that the distinction between MEIC's China-specific and global products was not sufficiently clarified. The MEIC team produces two distinct CO_2 emission products: a China-specific version (MEIC-China-CO_2) and a global version (MEIC-Global-CO_2). We selected the MEIC-Global-CO_2 product v1.0 based on its two primary advantages: it offers a higher spatial resolution (0.1°×0.1°) compared to the then-latest MEIC-China-CO_2 v1.4 (0.25°×0.25°), and its temporal coverage extends closer to the most recent years (1970–2023 vs 1970–2020). Importantly, while this product is globally scoped, the emissions calculation within the Chinese region retains the accuracy of a local inventory by using Chinese local energy statistics (from the China Energy Statistics Yearbook, CESY)) and emission factors (from the China Emission Accounts and Datasets, CEADs). We have revised content in Section 2.1, paragraph 3 to harmonize these descriptions and clarify that the global version was selected based on its superior technical specifications (spatial resolution and temporal coverage).

Revision:

Section 2.1, paragraph 3: *"…MEIC uses the transportation network data from the China Digital Road Network Map (CDRM) to constrain the distribution of vehicle activity as well as population density, GDP, and land use for other sectors (Li et al., 2017a; Xu et al., 2024b). <u>In this study, we</u>*

use the latest MEIC-Global-CO_2 product (v1.0), which provides higher spatial resolution (0.1° × 0.1°) and longer temporal coverage (1970-2023) than the MEIC-China-CO_2 product (v1.4; 0.25° × 0.25°, up to 2020). It's noteworthy that although MEIC-Global-CO_2 is a global product, its emissions calculations for China continue to rely on local energy statistics (CESY) and emission factors (CEADs), ensuring consistency with domestic data while improving spatiotemporal details."

2. Line 88: You mention standardising inventories to a common grid. Could this process introduce uncertainties? If so, could you quantify or acknowledge them?

Response: We appreciate the reviewer's comment. The standardization to a common grid introduces negligible additional uncertainty in this study. National and provincial totals are not affected, as they were derived either directly from the original inventory products or by spatially masking and summing emissions on their native grids.

For the gridded comparison, we adopted the MEIC grid as the spatial reference. For ODIAC, originally provided at a 1 km × 1 km resolution, emissions were spatially aggregated by summing all sub-grid values within each 0.1° × 0.1° cell to match the reference resolution. This aggregation preserves the total emissions without introducing interpolation-related errors. For EDGAR, CAMS, GEMS, and MEIC, all of which have the same resolution (0.1° × 0.1°) and identical latitude–longitude extents, we applied a nearest-neighbour method to ensure exact grid alignment. This approach maintains the original emission magnitudes and prevents artificial spatial gradients.

Therefore, the regridding and aggregation procedures do not substantially affect either the spatial distribution or the total emissions, and the associated uncertainties are considered negligible.

3. Line 174: Each growth phase is described with justification based on context, except from the third phase. Could you explain why emissions increase again after 2016?

Response: We thank the reviewer for this valuable suggestion. We have added an explanation for the renewed increase in CO_2 emissions after 2016. According to Zhang et al. (2020), the rebound was mainly driven by renewed infrastructure investment and the recovery of industrial activity after 2016. These developments substantially increased electricity demand, which was largely met by coal-fired power generation. As a result, fossil fuel consumption rose again, and the mitigation effect of the cleaner energy mix weakened compared with the 2012–2015 period. These points have been incorporated into the revised Results and Conclusion sections.

Revision:

(1) ***Section 3.1, paragraph 3:*** *"···From 2016 to 2023, all inventories show increased CO_2 emissions again, with a slower rate (0.30±0.016 Gt year-1) compared to the first phase. This rebound could be attributed to the expansion of infrastructure investment and the recovery of coal-based power generation, as the mitigation effect of the cleaner energy mix weakened after 2016 (Zhang et al., 2020)."*

(2) ***Section 4, paragraph 2:*** *"..., and a renewed growth of 0.30±0.016 Gt year^{-1} (2016–2023), mainly related to infrastructure-driven energy demand and coal use recovery following 2016. ..."*

Reference:

Zhang, Y., Zheng, X., Cai, W., Liu, Y., Luo, H., Guo, K., Bu, C., Li, J., and Wang, C.: Key drivers of the rebound trend of China's CO2 emissions, Environ. Res. Lett., 15, 104049, https://doi.org/10.1088/1748-9326/aba1bf, 2020.

4. Line 215: You explain spatial gaps in ODIAC and explain that they could be due to the inventory relying on night lightning. However, you do not mention other inventories. For example, are the spatial gaps in CAMS likely to be caused by similar reasons?

Response: We thank the reviewer for this insightful comment. We have examined the sectoral emissions of CAMS and found that the spatial gaps over western China mainly arise from the lack of aviation emissions. Specifically, CAMS includes only three transportation subsectors—Off-road transportation, Road transportation, and Ships—but does not account for aircraft emissions. To verify this, we compared the spatial distributions of transportation emissions among EDGAR, CAMS, MEIC, and GEMS (ODIAC does not provide sectoral data). As shown in the figure below, EDGAR, MEIC, and GEMS all display distinct emission patterns following major flight corridors over western China, while CAMS shows only the road transport pattern. This confirms that the absence of aviation emissions in CAMS explains the spatial gaps observed in that region. We have revised the content in Section 3.2.1, paragraph 1 to explain the spatial gap in CAMS.

Revision:

Section 3.2.1, paragraph 1: *"…, while regions with limited nighttime lighting, including both sparsely populated areas and areas with high population but limited lighting, such as Western Sichuan, Inner Mongolia, and Xinjiang, are not captured. <u>By contrast, the spatial gaps over western China in CAMS (Fig. 4d) mainly arise from the lack of aviation emissions. CAMS accounts for transport emissions from road, off-road, and ships but omits aviation. As shown in Figure S1, EDGAR, MEIC, and GEMS capture distinct emission bands along major flight corridors over western China, whereas CAMS only shows the road transport pattern, explaining the missing emissions over western China.</u>"*

Section 7, Figure S1:

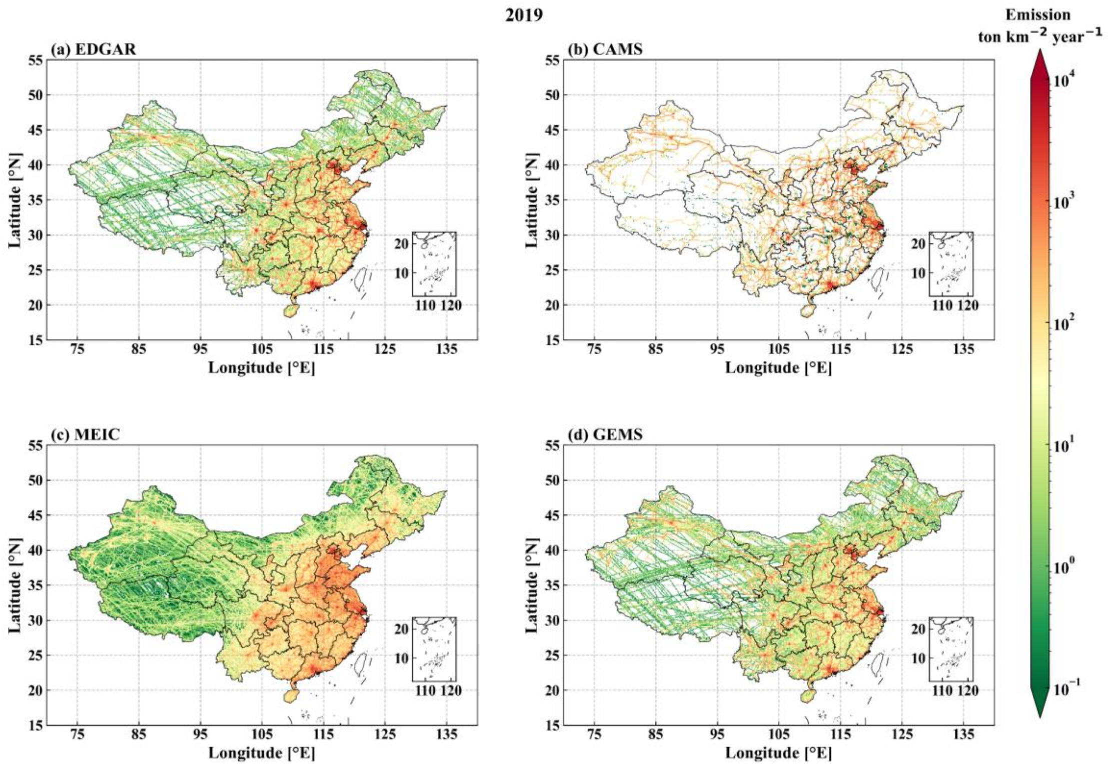
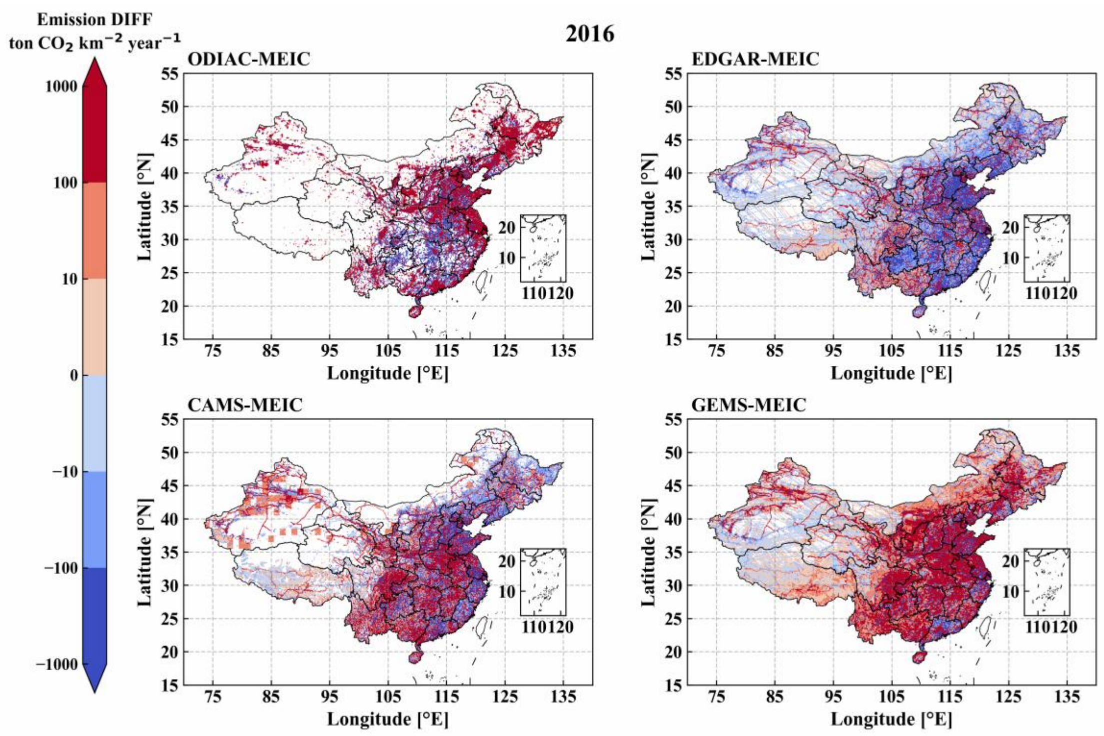


Figure S1. *Spatial distribution of CO_2 emissions from transport sector in 2019 across four inventories (EDGAR, CAMS, MEIC, and GEMS).*

5. Figure 4: Why was 2019 chosen as the base year? Would spatial patterns differ significantly in other years?

Response: We thank the reviewer for the comment. As mentioned in Section 3.2.1, 2019 was chosen as the reference year because it is the most recent year for which all five gridded inventories (ODIAC, EDGAR, MEIC, CAMS, and GEMS) provide spatially explicit emission data. Moreover, 2019 represents a typical pre-pandemic year, unaffected by the COVID-19 lockdowns in 2020-2021 2019 is free from exceptional events such as the COVID-19 lockdowns, making it a representative baseline for comparison.

Although our manuscript focuses on 2019 due to space limitations, we also conducted preliminary analyses for the third emission phase (2016-2023). As illustrated in the GIF below, the spatial patterns of inter-inventory differences remain generally consistent over time, although the overall magnitude of emissions varying. The only notable exception occurs in the EDGAR–MEIC comparison, where differences in southwestern China shift from obvious positive to negative during 2016–2017. After 2017, the EDGAR–MEIC spatial differences stabilize, and other inventories relative to MEIC show minimal spatial variation throughout 2016–2023.



Temporal evolution of spatial differences in CO_2 emissions between MEIC and other inventories (ODIAC, EDGAR, CAMS, and GEMS) during 2016–2023.

6. Figure 5: Inventories are compared to MEIC as a baseline. Could you comment on the existing uncertainties relating to MEIC, and what this means for the results?

Response: We thank the reviewer for this valuable comment. We acknowledge that MEIC itself is subject to uncertainties, mainly arising from the underlying activity data, emission factors, and spatial proxy selection. However, MEIC remains one of the most recognized and used when studying anthropogenic emissions in China. For example, it has been integrated into the MIX inventory as the Chinese component of the Asian anthropogenic emissions (Li et al., 2017) and was used to develop high-resolution (1 km × 1 km) emission maps for 2013 (Zheng et al., 2021). Previous studies have also shown that simulations based on MEIC are more consistent with observations than those using EDGAR or ODIAC in Beijing (Che et al., 2022) and perform better in Xianghe and Xinlong (Yang et al., 2025). Therefore, we think MEIC can serve as a reasonable benchmark for spatial comparison. Nevertheless, the uncertainties in MEIC imply that our spatial difference maps (Fig. 5) reflect relative differences among inventories rather than absolute errors. We have added this clarification to the revised manuscript.

Revision:

Section 3.2.1, paragraph 3: *"To assess spatial consistency, we compared ODIAC, EDGAR, CAMS, and GEMS with MEIC as a benchmark (Fig. 5). <u>MEIC was chosen because it is compiled using local statistics and has been widely applied and validated in previous studies (Li et al., 2017b; Zheng et al., 2021; Che et al., 2022; Yang et al., 2025), making it a reasonable reference for comparison.</u> ..."*

References:

Che, K., Cai, Z., Liu, Y., Wu, L., Yang, D., Chen, Y., Meng, X., Zhou, M., Wang, J., Yao, L., and Wang, P.: Lagrangian inversion of anthropogenic CO_2 emissions from Beijing using differential column measurements, Environ. Res. Lett., 17, 075001, https://doi.org/10.1088/1748-9326/ac7477, 2022.

Li, M., Zhang, Q., Kurokawa, J., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmos. Chem. Phys., 17, 935–963, https://doi.org/10.5194/acp-17-935-2017, 2017.

Yang, H., Wu, K., Wang, T., Wang, P., and Zhou, M.: Atmospheric anthropogenic CO2 variations observed by tower in-situ measurements and simulated by the STILT model in the Beijing megacity region, Atmospheric Research, 325, 108258, https://doi.org/10.1016/j.atmosres.2025.108258, 2025.

Zheng, B., Cheng, J., Geng, G., Wang, X., Li, M., Shi, Q., Qi, J., Lei, Y., Zhang, Q., and He, K.: Mapping anthropogenic emissions in China at 1 km spatial resolution and its application in air quality modeling, Science Bulletin, 66, 612–620, https://doi.org/10.1016/j.scib.2020.12.008, 2021.

7. Lines 235–240: Inventory users would benefit from specific interpretation for all inventories. For example, why does ODIAC allocate more emissions to areas? Is it related to night lighting again? Why is the CAMS pattern opposite to ODIAC?

Response: We thank the reviewer for this valuable comment. The emission hotspots in ODIAC are more concentrated in regions with intense nighttime lighting because ODIAC uses satellite nightlight data as area source proxy for fossil fuel emissions. It's also important to clarify that the opposite spatial patterns between CAMS and ODIAC refer to the spatial differences relative to MEIC (i.e., inventory minus MEIC), rather than their direct emission distributions. The discrepancies mainly arise from different spatial allocation methods adopted by each inventory. ODIAC allocates fossil fuel emissions based on satellite nightlight as area source, leading to higher emissions in coastal and urbanized regions where nightlight signals are strong. In contrast, CAMS dose not rely on nightlight data. It builds upon EDGAR and CAMS-GLOB-Ship for its spatial distribution, which may tend to assign relatively more emissions to inland regions. Consequently, when compared with MEIC, ODIAC shows higher emissions along the eastern coast, while CAMS displays higher values over several inland provinces, producing opposite spatial difference patterns.

8. Lines 241- 248: You explain that EDGAR has very large extremes in 0.14% of grid cells, likely due to EDGAR allocating emissions aggressively to point sources (and using outdated CARMA). The presence of such large extremes, which strongly influence averages, raises questions about the robustness of EDGAR as an inventory. Should this be a concern for users? Could you clarify how using MEIC as a baseline may influence this result?

Response: We thank the reviewer for highlighting this important point regarding the relatively extreme values observed in the EDGAR inventory and their implications. At present, we don't have independent observations to evaluate whether these extreme values are accurate. Based on our analysis, these extremes most likely originate from the sector industry and construction, due to EDGAR's use of point-source information from CARMA, which may not accurately capture the spatial distribution and emission magnitudes of power plants in China.

Given the presence elative extreme values, we suggest that users exercise caution when using EDGAR to study the spatial distribution of emissions from the sector industry and construction or power plants in China. Specifically, users should compare EDGAR with multiple inventories, conducting cross-inventory analyses to ensure robust interpretations of spatial patterns.

Regarding the use of MEIC as a baseline, it is reasonable for China-focused studies because it is compiled using Chinese statistical data and has been widely applied and validated in previous studies (Li et al., 2017; Zheng et al., 2021; Che et al., 2022; Yang et al., 2025). MEIC provides a locally informed reference that allows identification of relative differences between EDGAR and locally detailed inventories.

We have emphasized these points in the Conclusion and Discussion section of the manuscript, explicitly recommending that future analyses account for the relatively extreme values in EDGAR and validate results using multiple inventories wherever possible.

Revision:

Section 4, paragraph 3: *"…The ODIAC nightlight proxy distributes more emissions in urban areas and fewer emissions in the western regions. EDGAR, which is based on the CARMA database, concentrated power plant emissions on fewer grids, resulting in extreme anomalies where the difference (EDGAR-MEIC) exceeds 10^5 ton CO_2 km^{-2} $year^{-1}$. <u>These high-value grids underscore the importance of cross-inventory comparisons when using EDGAR to analyze the spatial distribution of industry sector or power plant emissions in China.</u> In contrast, MEIC uses the more detailed CPED and distributes similar total CO_2 emissions…"*

References:

Che, K., Cai, Z., Liu, Y., Wu, L., Yang, D., Chen, Y., Meng, X., Zhou, M., Wang, J., Yao, L., and Wang, P.: Lagrangian inversion of anthropogenic CO $_2$ emissions from Beijing using differential column measurements, Environ. Res. Lett., 17, 075001, https://doi.org/10.1088/1748-9326/ac7477, 2022.

Li, M., Zhang, Q., Kurokawa, J., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmos. Chem. Phys., 17, 935–963, https://doi.org/10.5194/acp-17-935-2017, 2017.

Yang, H., Wu, K., Wang, T., Wang, P., and Zhou, M.: Atmospheric anthropogenic CO2 variations observed by tower in-situ measurements and simulated by the STILT model in the Beijing megacity region, Atmospheric Research, 325, 108258, https://doi.org/10.1016/j.atmosres.2025.108258, 2025.

Zheng, B., Cheng, J., Geng, G., Wang, X., Li, M., Shi, Q., Qi, J., Lei, Y., Zhang, Q., and He, K.: Mapping anthropogenic emissions in China at 1 km spatial resolution and its application in air quality modeling, Science Bulletin, 66, 612–620, https://doi.org/10.1016/j.scib.2020.12.008, 2021.

9. Line 299: You find large difference between CEADS (provinces) and CEADS (sectors) for Shanxi. You conclude that sector-level estimates should be prioritised, as the sum of all provinces estimates do no match the national estimates. Could you comment on why province-level estimates are so uncertain, and different from sectors estimates?

Response: We thank the reviewer for this insightful comment. We examined CO_2 emissions from CEADs (sectors) and CEADs (provinces) for Shanxi and found that the large discrepancy mainly originates from raw coal-related emissions, which is the dominant contributor to total emissions (Wei, 2022). As shown in the figure below, CO_2 emissions from raw coal in CEADs (provinces) are on average 665 Mt year^{-1} higher than those in CEADs (sectors), resulting in an overall mean difference of 512 Mt year^{-1} between the two datasets. This indicates that inconsistencies in fuel-specific accounting, particularly for raw coal, are a key contributor to the provincial-level uncertainty. The detailed comparison has been added to the Supplementary Material, and we have clarified this in the revised manuscript.

Revision:
Section 3.3.1, paragraph 1: *"…In contrast, the CEADs (sectors) closely matches the other five independent inventories (ODIAC, EDGAR, MEIC, CAMS and GEMS), with its mean emissions deviating by no more than 3.84 Mt year^{-1} from the average of the five inventories. The large discrepancy between CEADs (provinces) and CEADs (sectors) mainly originates from the much higher raw coal–related emissions in CEADs (provinces) (Fig. S3), as coal is the dominant contributor to total emissions (Wei, 2022)."*

Section 7, Figure S3:

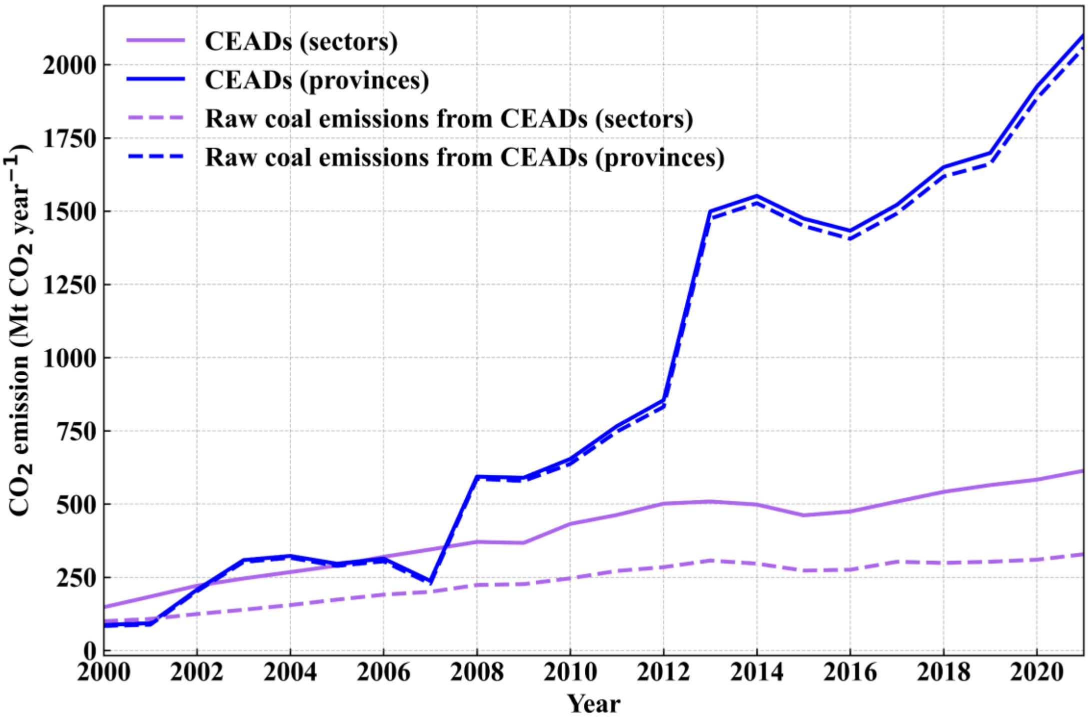


Figure S3. *Comparison of total CO_2 emissions and raw coal–related CO_2 emissions in Shanxi from CEADs (sectors) and CEADs (provinces) during 2000–2020. Solid lines represent total emissions, while dashed lines indicate emissions from raw coal combustion.*

Reference:

Wei, C.: Historical trend and drivers of China's CO2 emissions from 2000 to 2020, Environ Dev Sustain, 1–20, https://doi.org/10.1007/s10668-022-02811-8, 2022.

10. Section 3.3.2: In this section, you analyse timeseries for nine specific provinces, chosen based on a classification. This section currently reads as a descriptive list without a clear narrative or takeaway. Could you clarify the aim e.g., to illustrate provincial heterogeneity between inventories.

Response: Thank you for this comment. The criteria for selecting the nine representative provinces and research objectives in Section 3.2.2 may not have been clearly stated. The main purpose of this analysis was to investigate how inventory consistency and discrepancies vary between provinces with higher total emissions and those with high inter-inventory uncertainty over the 2000–2023 period.

To identify these representative provinces, we ranked all provinces each year in descending order by (1) their six-inventory mean CO_2 emissions and (2) their six-inventory standard deviation (SD). Each province therefore obtained an annual rank in both metrics for each year (2000–2023). The cumulative rank score was then calculated as the sum of annual ranks across all years, representing the long-term magnitude or variability of emissions. Provinces with the smallest cumulative rank scores in each category were selected as representative provinces (Table 3). This approach highlights provinces that consistently contribute the most to national totals and those that exhibit the largest inventory discrepancies throughout the study period. To improve clarity, we have revised Section 3.2.2 to more explicitly describe the purpose of this analysis and the selection procedure, as shown below.

Revision:
Section 3.3.2, paragraph 1: *"The mean and SD of the provincial CO_2 emissions from 2000 to 2023 are shown in Figure S4. To investigate how inter-inventory consistency and discrepancies vary across provinces with high emissions or uncertainties, we select a subset of representative provinces for a detailed comparison. Representative provinces are identified using the SD and the mean emissions between the six emission inventories, calculated for the period 2000-2023. Each year, all provinces are ranked in descending order based on these two metrics. The cumulative scores are calculated by summing the annual ranks over the entire 24-year period (2000-2023), reflecting each province's long-term ranking in terms of emission magnitude or SD. A lower cumulative score indicates higher mean emissions or emission uncertainties (SD). The top six provinces in each category are selected, resulting in a list of nine representative provinces"*

11. Line 361: You state that results are opposite to Han et al. (2020b). Could you give more details about these differences? Why do different versions of inventories give such different results? What does this mean for inventory users?

Response: We thank the reviewer for this insightful question. The opposite results relative to Han et al. (2020b) mainly arise from differences in the inventory versions used. Han et al. (2020b) employed EDGAR v4.3.2 and MEIC v1.3, while our study uses the most recent versions, EDGAR 2024 and MEIC-Global-CO_2 v1.0. To examine this discrepancy, we compared the national totals of these four datasets, as shown in the figure below. The two EDGAR versions show nearly identical emission trends between 2000 and 2012, with EDGAR 2024 being only slightly higher (0.00085 Gt year^{-1}). In contrast, MEIC v1.3 reports substantially higher national emissions (by about 1.43 Gt year^{-1}) than MEIC-Global-CO_2 v1.0 during 2008–2017, and its estimates are close to those of EDGAR 2024 (difference of 0.30 Gt year^{-1}).
These findings indicate that the divergence primarily results from the updated MEIC version, which yields lower national totals than its earlier release. However, the MEIC database does not provide detailed documentation on version updates, limiting our ability to trace the exact methodological changes. This underscores the importance for inventory users to carefully consider version differences when conducting trend analyses or cross-inventory comparisons.

Revision:
Section 4, paragraph 1: *"China's annual anthropogenic CO_2 total emission increases from 3.42 Gt in 2000 to 12.03 Gt in 2023. When compared with the officially reported NGHGI and the six-inventory mean, CAMS shows the smallest deviation from the NGHGI, while ODIAC agrees most closely with the multi-inventory mean. The six inventories display a broadly consistent emission trend, but their discrepancies among the inventories have widened from 0.41 Gt year^{-1} to 1.63 Gt year^{-1}, mainly due to the highest estimates reported from EDGAR and the lowest values estimated*

from MEIC, especially after 2012. Our results are consistent with Zheng et al. (2025) but opposite to Han et al. (2020b), demonstrating the differences in emission versions (Our study: EDGAR2024, MEIC-global-CO_2 v1.0; Zheng: EDGAR v7.0, MEIC-China-CO_2 v1.4; Han: EDGAR v4.3.2, MEIC-China-CO_2 v1.3). A comparison between these versions (Fig. S6) shows that the divergence mainly arises from a downward revision in the latest MEIC dataset, which reports about 1.43 Gt year^{-1} lower emissions on average over 2008–2017. In contrast, EDGAR's national totals remained nearly unchanged across versions, with differences within 0.001 Gt year^{-1} during 2000-2012. These results highlight the significant impact of inventory version updates on comparative emission analyses."

Section 7, Figure S6:

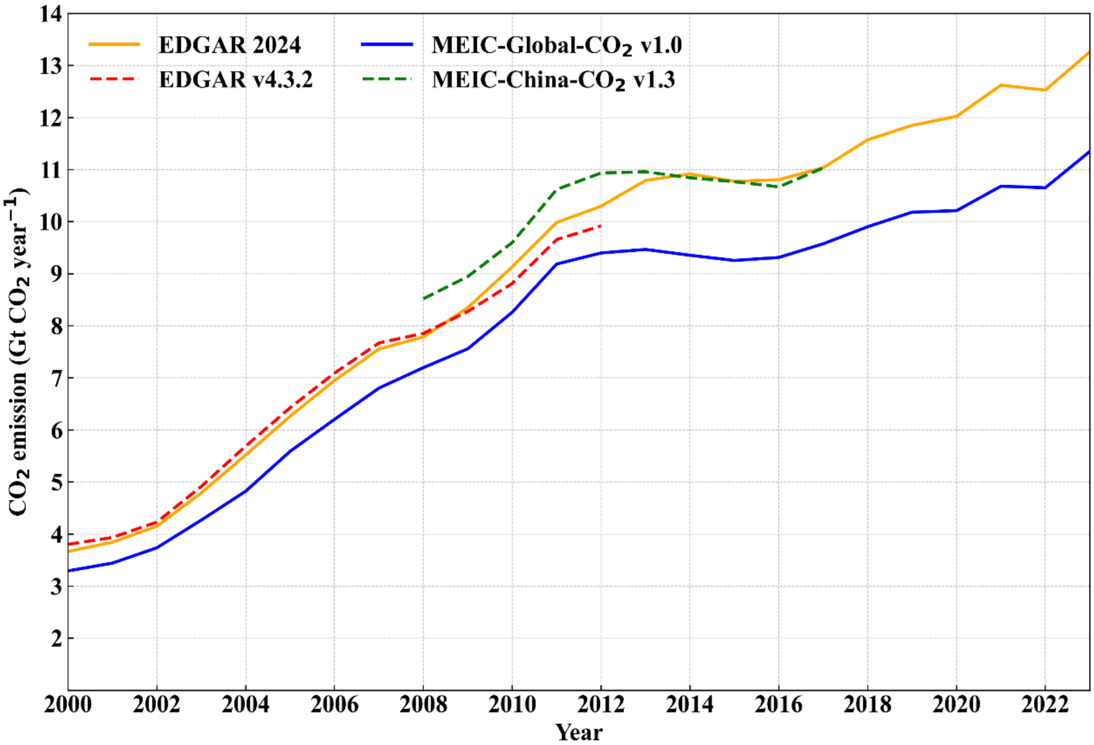


__Figure S6.__ Comparison of national CO_2 emissions from different versions of the EDGAR and MEIC inventories. The older versions (EDGAR v4.3.2 and MEIC-China-CO_2 v1.3) used in Han et al. (2020b) are compared with the updated versions (EDGAR 2024 and MEIC-Global-CO_2 v1.0) used in this study.

12. Line 388: "Ensemble approaches" please define how this method would be used and explain why they would help mitigate biases

Response: We thank the reviewer for this insightful comment. The term ensemble approaches in this study refers to statistical and model-based frameworks that integrate multiple emission inventories and auxiliary datasets (e.g., energy statistics, spatial proxies, and inversion-based flux estimates) to produce a consensus estimate and quantify uncertainty. Such approaches can take various forms, including weighted averaging, Bayesian inversion, or ensemble learning in machine learning. By combining independent datasets with different methodological assumptions and spatial representations, ensemble techniques reduce the influence of biases or errors present in any single inventory and provide more robust emission estimates.