Reviewer #2

✧ **General comments:**

This manuscript presents a novel approach to improving the annual mercury emissions inventory for China from 1978 to 2021 using the P-CAME model. The work is important for understanding mercury emissions in the region and for supporting policy measures under the Minamata Convention. However, the validation section requires significant improvement to ensure the reliability of the data and the robustness of the conclusions. Below are some suggestions to enhance the manuscript before it can be considered for publication.

Response: We appreciate the reviewer's comments and have addressed them in our revised manuscript according to the detailed comments. Thank you for the comments!

✧ **Detailed comments and replies**

1. Validation over the Entire Study Period: The study covers a long time span (1978–2021), with peak emissions identified around 2010–2012, as shown in Figure 3. However, the model evaluation is limited to the year 2021, which represents a period of reduced emissions compared to the peak years. This raises concerns about whether the model performs well in earlier years, especially around the time of peak emissions. To address this issue, the authors should include validation for multiple years, particularly around periods of significant changes in emissions, such as 2010–2012. If observational data from earlier periods are scarce, the authors could explore alternative methods to compare model outputs with historical trends.

Response: We appreciate the reviewer's comments. We added long-term Hg^0 concentration simulation and comparison with observation data at 9 sites during 2009-2021. Among these sites, 6 of them included observation Hg^0 concentration data before 2012, which provides us chances to observe the impacts of peak emissions.

"3.5 Long-term simulation of atmospheric mercury concentrations

The temporal and spatial distributions of annual atmospheric Hg^0 concentration are presented in Fig. 5. During 2011 to 2021, the simulated Hg^0 concentrations showed a declining trend, with the maximum values decreasing from 5.7 ng/m^3 to 3.0 ng/m^3, and the national average dropping slightly from 1.5 ng/m^3 to 1.4 ng/m^3. The spatial distribution analysis (Fig. 5) highlights a decline of simulated Hg^0 concentration in high-emission regions. However, the simulated magnitude of decline fails to capture the observed decline at monitoring sites, primarily due to an underestimation of Hg^0 concentrations from 2010-2013, when anthropogenic emissions peaked in China (Fig. S5). This issue has also been existed in previous studies, which found that GEOS-Chem simulations underestimate Hg^0 concentration during this period (Liu et al., 2019; Sun et al., 2024). The underestimation may stem from either the model or our anthropogenic emission inventory. Observational studies have shown that the decline in anthropogenic emissions is the key driver behind the decrease in Hg^0 concentrations at both background sites (Changbai, Ailao, Damei, Waliguan, Chongming) (Feng et al., 2024; Tang et al., 2018), and urban sites (Nanjing) (Sun et al., 2024). To explore reasons for simulation underestimation, we compared the decline rates of observed Hg^0 concentration, simulated Hg^0 concentration and anthropogenic emissions at these sites, as shown in Table 1. For each site, the decline rate of observed Hg^0 concentration was calculated as the difference between maximum value and the concentration at the end of observation period, divided by the

maximum value (see Equation S9 for an example calculating at Changbai). The same method was applied to calculate decline rates for simulated Hg^0 concentration, national total Hg^0 emissions, Hg^0 emissions from the 9 surrounding grids (approximately 500 km × 500 km), and Hg^0 emissions from the current grid over the same period.

As shown in Table 1, the decline rates of observed Hg^0 concentrations vary across different site types based on their location and emission impacts: (1) Background sites (Changbai, Ailao, Damei, Waliguan): These high-altitude sites with minimal local emissions represent national even global impacts. Their observed Hg^0 concentration decline rates closely align with the national total Hg^0 emission decline rates and are significantly higher than simulated Hg^0 concentration decline rates. (2) Regional background sites (Chongming, Miyun): Located in suburban areas, these sites reflect regional impacts. Their observed Hg^0 concentration decline rates align more closely with the emission decline rates from nearby grids (9 surrounding grids) and are also much higher than simulated Hg^0 decline rates. (3) Urban sites (Nanjing, Tsinghua, Hohhot): Urban sites are influenced by diverse emission sources, making it difficult to directly associate observed Hg^0 concentrations with specific emission types. At Nanjing site, impacted by point source emissions from CFPP and CEM within the local grid, the observed decline rates closely align with local emission decline rates and are higher than simulated rates. At Tsinghua site, impacted by transported emissions from adjacent provinces, the observed Hg^0 decline rates are comparable to the national total Hg^0 emission decline rates. At Hohhot site, situated at a high altitude and impacted by broader area emissions, the observed Hg^0 decline rates align with national total Hg^0 emission decline rates.

The observed decline rate matches the emission decline rate and exceeds the simulated rate at all sites. This suggests that our anthropogenic emissions inventory is reasonable and should have reproduced the observed trends. Potential reasons for the model's underestimation include: (1) Boundary conditions. Boundary conditions play a critical role in determining the global background concentration of Hg^0 in nested simulations. However, global anthropogenic emissions used in simulations often fail to capture the observed decline trend in Hg^0 concentrations. For example, observations from the Northern Hemisphere indicate a decline of approximately 0.011 ng m-3 yr-1, while simulations show only a slight decline of 0.0014 ng m-3 yr-1 (Feinberg et al., 2024). This discrepancy introduces bias in nested simulation trends, particularly at background sites. The inability of boundary conditions to reflect observed trends highlights a key limitation in current simulation. (2) Legacy re-emissions. Legacy re-emissions refer to the re-emission of previously deposited Hg. These Hg^0 emissions diffuse back into the atmosphere and are reported to contribute significantly to current atmospheric mercury concentration (Angot et al., 2021) or deposition (Amos et al., 2013). For example, studies suggest that legacy re-emissions account for approximately 60% of atmospheric deposition, compared to 27% from anthropogenic emissions (Amos et al., 2013). (3) Transport process and wind field. Transport process plays a critical role in controlling Hg^0 concentrations and trends (Roy et al., 2023), with wind field being a key factor in determining transport process (Brasseur and Jacob, 2017; Yang et al., 2024). By comparing simulated 10 m wind speed from MERRA2 with observed wind speed, we found discrepancies in the monthly wind speed trends between MERRA2 and meteorological observations (Fig. S6). These inconsistencies in monthly trends suggest a potential bias in MERRA2 wind speed data, consistent with findings from other evaluation studies (Miao et al., 2020). Similar biases are observed in wind direction when

comparing MERRA2 with observations (Fig. S7). These biases likely contribute to transport simulation errors and may significantly underestimate Hg^0 concentrations in the model.

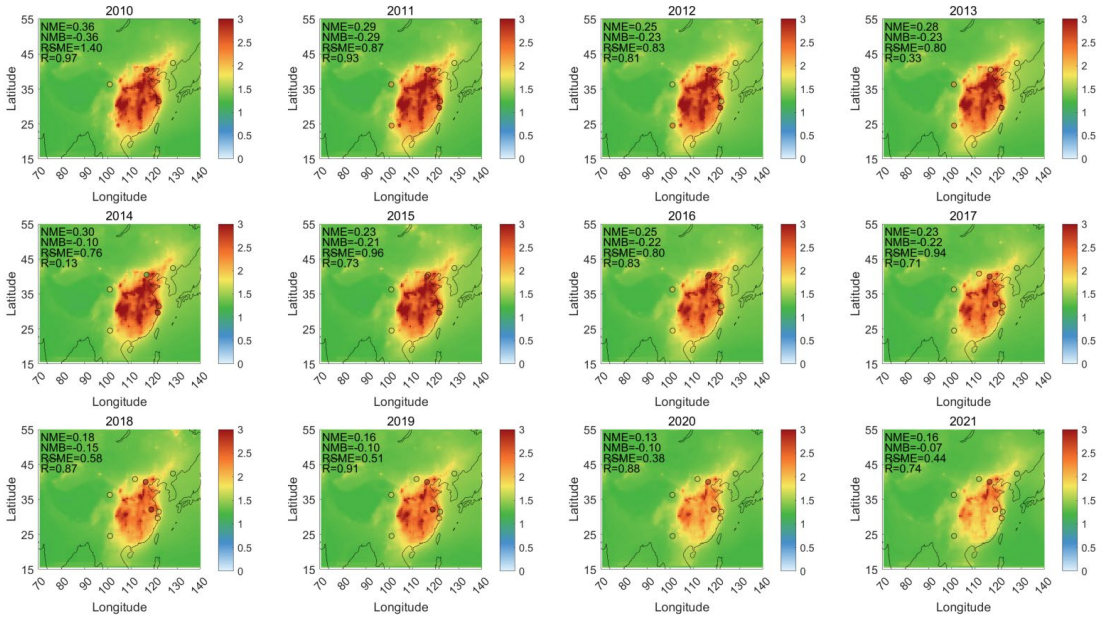

Figure 5 Temporal and spatial distribution of simulated Hg^0 concentration (ng/m^3).

Table 1 Decline rate of observed Hg^0 concentration, Hg^0 emissions, and simulated Hg^0 concentration

Sites	Altitude (m a.s.l.)	Type	Period	Observed Hg^0 concentration	Simulated Hg^0 concentration	Decline rate		
						National total Hg^0 emissions	Hg^0 emission of surrounding 9 grids	Hg^0 emission of current grid
Changbai	741	Background	2013-2021	0.22	0.04	0.30	0.58	0.58
Ailao	2450	Background	2012-2021	0.42	0.03	0.35	0.12	0.09
Damie	550	Background	2012-2021	0.46	0.25	0.35	0.38	-0.14
Waliguan	3816	Background	2013-2021	0.29	-0.02	0.30	0.13	-0.12
Chongming	10	Regional Background	2010-2021	0.46	0.21	0.36	0.69	-0.36
Miyun	128	Regional Background	2010-2016	0.31	0.08	0.17	0.42	0.36
Nanjing	10	Urban	2017-2021	0.37	0.19	0.15	0.17	0.35
Tsinghua	50	Urban	2015-2021	0.32	0.06	0.26	0.29	0.38
Hohhot	1100	Urban	2017-2021	0.32	0.05	0.15	0.15	0.04

”

See revised Manuscript, Lines 308-361.

2. Extended Validation Metrics: While the model evaluation provides normalized mean bias (NMB) and normalized mean error (NME) for 2021, these metrics alone may not fully capture the model's performance. I suggest incorporating additional metrics, such as the root mean square error (RMSE) and the correlation coefficient (R), to provide a more comprehensive evaluation. Moreover, the evaluation should be conducted seasonally and include spatial analysis to account for variations in mercury emissions throughout the year. This will help ensure the model's performance across different geographic regions and emission sources.

Response: We appreciate the reviewer's comments. We have added RMSE and R to evaluate model performance for both only proxy-based and P-CAME at 9 sites in 2020, as this year shows less bias based on the spatial evaluation (Figure 5). Seasonal simulations and observations were compared, and corresponding NMB, NME, RMSE, and R were calculated for each site, as shown in Figure 6.

"3.6 Simulation comparison using P-CAME and only proxy-based inventory

We selected 2020 to compare the simulation differences between the P-CAME and only proxy-based inventories, as 2020 exhibits less bias according to Fig. 5. For each site, we compared seasonal average Hg^0 concentrations and evaluated performance using NMB, NME, RMSE, and R, as detailed in Fig. 6. Our analysis revealed that P-CAME have the potentiality to improve simulation accuracy for urban sites, such as Nanjing and Hohhot. In Nanjing site, the grid containing the Nanjing site includes CFPP and CEM point sources. The only proxy-based method underestimates emissions compared to P-CAME (Fig. S8), resulting in lower simulated Hg^0 concentrations. P-CAME reduces simulation bias, yielding lower NMB, NME, and RMSE values, indicating better agreement with observations. In Hohhot site, the only proxy-based method tends to overestimate emissions due to the high population density (Fig. S8). By contrast, P-CAME produces lower simulated Hg^0 concentrations, which better align with observations, with lower NMB, NME, and RMSE values. These two sites highlight two common scenarios: (1) overestimated emissions in densely populated areas and (2) underestimated emissions in industrial clusters, as discussed in Section 3.1. From this perspective, P-CAME has the capacity to reduce simulation bias by more accurately allocating spatial emissions in urban regions. However, this capacity is currently limited by model bias, such as poor performance in simulating transport processes, as discussed in Section 3.5. For urban sites like Qingdao and Tsinghua, seasonal trends are influenced by air mass sources from different directions, driven by air pressure changes between land and ocean (Shao et al., 2022; Wang et al., 2021). For example, we found that the wind field from MERRA2 does not closely match observations (Fig. S7), which could lead to simulation bias. Since the model struggles to accurately capture these transport processes, its performance at these sites is poor, making it more challenging to identify improvements from revising the emission inventory. The model performs relatively better at rural sites when compared with observations. At these locations, there is little difference in simulation outcomes between using P-CAME and the only proxy-based inventory.

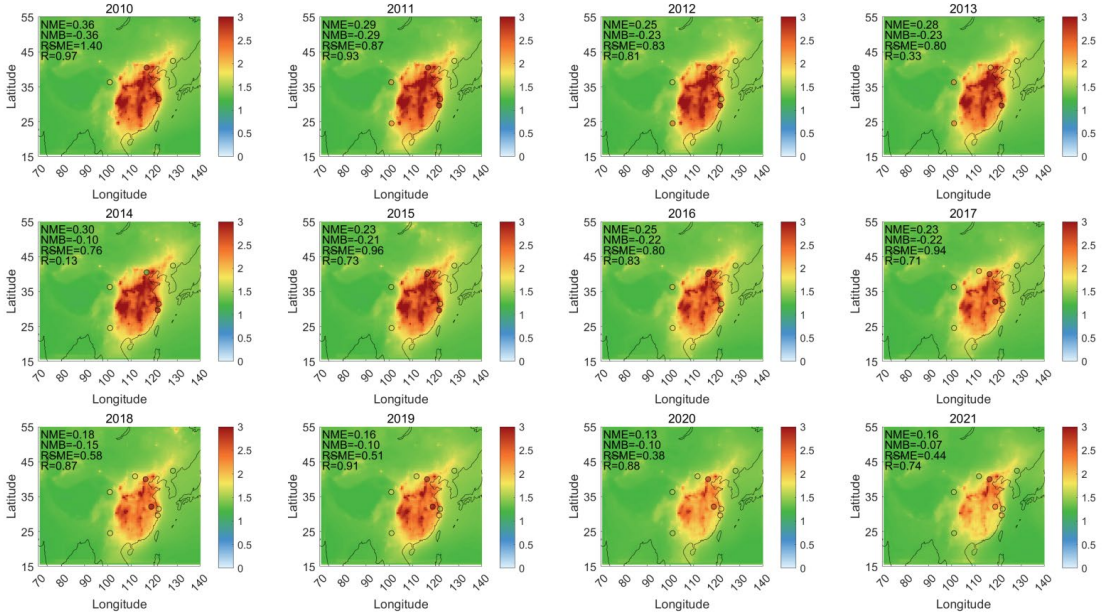

Figure 5 Temporal and spatial distribution of simulated Hg^0 concentration (ng/m^3).

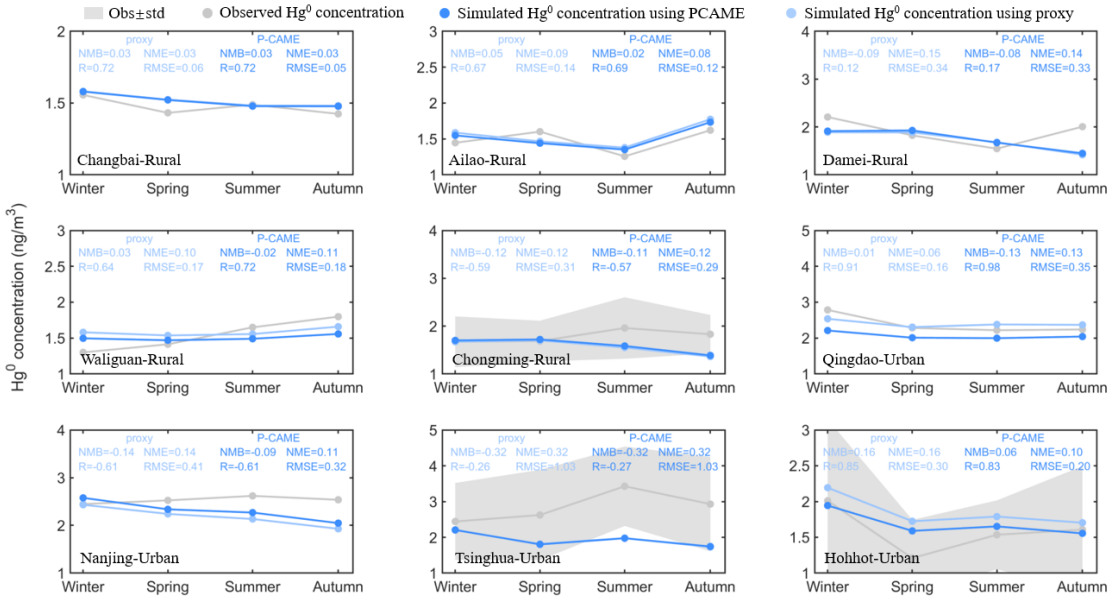

Figure 6 Comparison of observed and simulated atmospheric mercury concentrations using only proxy-based and P-CAME inventory.

”

See revised Manuscript, Lines 358-359 and Lines 362-384.

3. Definition of GEM (L 247): On line 247, the manuscript introduces "GEM" without defining it. While GEM is a well-known term in mercury studies, it is important to spell out "Gaseous Elemental Mercury (GEM)" upon first use to ensure clarity, especially for readers less familiar with the topic.

Response: Thank you for your suggestion. We have replaced all instances of GEM with Hg^0

and defined Hg^0 (Gaseous elemental mercury) upon its first use.

"P-CAME also demonstrates consistency with observed gaseous elemental mercury (Hg^0) concentration trends over the past decade"

See revised Manuscript, Lines 28-29.

4. Improving Figure 4: In Figure 4, the authors use a bar plot to compare observed and modeled GEM concentrations. While this provides some insight, a range plot (mean with standard deviation) or a box-and-whisker plot would be a better way to represent the variability in the data. Furthermore, a scatter plot could be added to show the correlation between observed and modeled data points, helping readers assess whether the model accurately captures the distribution of GEM concentrations.

Response: Thank you for your suggestion. We have replaced the bar plot with a range plot and included NMB, NME, RMSE, and R in the figure to better illustrate the data variability.

"

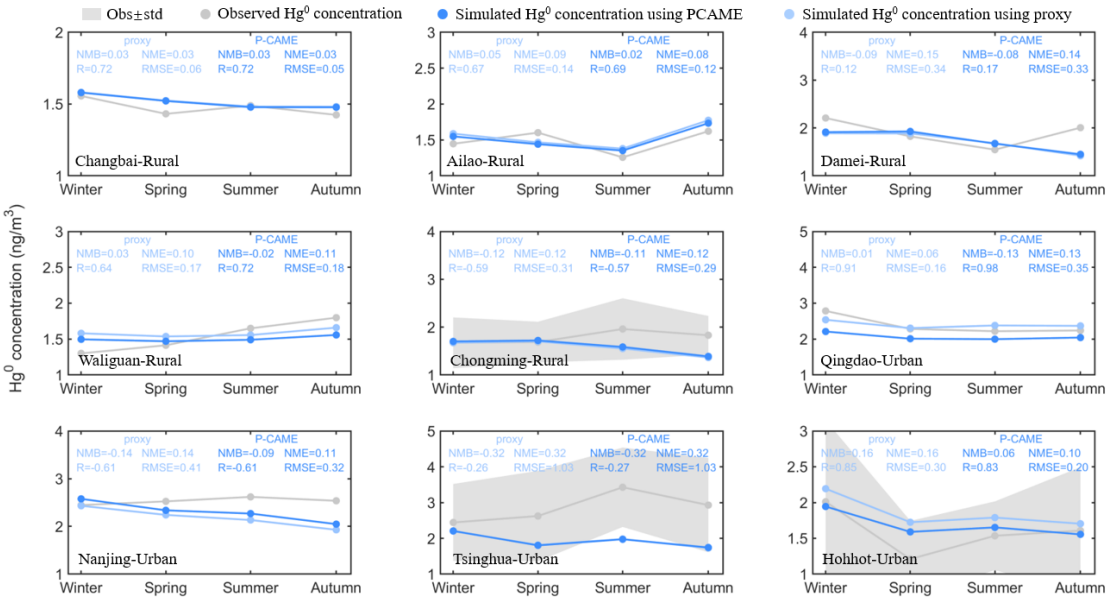

Figure 6 Comparison of observed and simulated atmospheric mercury concentrations using only proxy-based and P-CAME inventory.

"

See revised Manuscript, Lines 382-384.

5. Clarification of Data Availability (L 294): The manuscript mentions the availability of annual mercury emission inventories on figshare. However, the data in figshare (1978, 1980, 1985, etc.) do not appear to match the continuous data shown in Figure 3. It is important to clarify whether "all annual data from 1978 to 2021" will be made publicly available. If only certain years will be shared, this should be clearly stated in both the manuscript and the data repository to avoid confusion.

Response: Thank you for your comment. We have uploaded all annual data from 1978 to 2021 to figshare.

"Integrating point source emission inventory (P-CAME) can be accessed from http://doi.org/10.6084/m9.figshare.26076907 (Cui et al., 2024)."

See revised Manuscript, Lines 386-387 and figshare link.