Below, we provide a detailed, point-by-point response to the reviewer's comments and describe the corresponding revisions made to the manuscript. Reviewer comments are shown in pink, author responses and original manuscript text are shown in black, and revised manuscript text is shown in blue.

This manuscript addresses an important and timely issue: mercury bioaccumulation models are often too simplified to reproduce observed responses to mercury emission reduction strategies and tend to predict overly linear outcomes compared to empirical observations. In my view, this represents a meaningful contribution to modelling efforts supporting the Minamata Convention. However, the current version lacks a mechanistic discussion explaining why the assembled models behave in this linear manner. The authors have done a great job compiling the existing model results, and with additional depth in the discussion, the paper could not only identify structural weaknesses in current models but also highlight lessons for improving future model development.

Additionally, there are several unclear sections, especially in the methods, that could be further clarified. Therefore, I recommend acceptance with major revisions.

Author response

We sincerely thank the reviewer for this thoughtful and constructive assessment of our manuscript. We greatly appreciate the reviewer's recognition that the study addresses an important and timely question and has the potential to contribute meaningfully to mercury modeling efforts relevant to the Minamata Convention.

We agree that the previous version of the manuscript would benefit from a clearer mechanistic explanation of the relatively linear responses produced by many of the assembled modeling studies. In response, we have added targeted revisions to the Discussion (Section 3.3, Pages 24–26) to more explicitly explain how these patterns may arise from shared structural assumptions across current modeling frameworks. In particular, we clarify that many models implicitly or explicitly assume (i) atmospheric deposition as the dominant driver of future fish MeHg change, (ii) relatively direct and rapid propagation of changing Hg inputs through watershed and aquatic compartments, and (iii) limited buffering from long-term legacy Hg stores in soils, wetlands, and sediments. We also note that simplified representations of watershed retention, biogeochemical

transformation, and food-web dynamics may contribute to the comparatively linear responses observed in model projections relative to empirical observations.

We also appreciate the reviewer's comments regarding clarity, particularly in the Methods section. In response, we have made targeted revisions to improve transparency and reproducibility. Specifically, on Page 4, we clarified the computational environment by specifying the Python version (Python 3.12.13, Google Colab) and the packages used (scikit-learn, pandas, NumPy, SciPy, and matplotlib). In addition, on Pages 18–19, we added explicit equations describing how scenario-based percentage and annual changes in Hg emissions, deposition, and fish MeHg were calculated.

We are grateful for these helpful suggestions, which have improved the clarity and framing of the revised manuscript.

Specific comments

- Line 26: The sentence's mentioned Hg is identified by the WHO as a top 10 chemical but doesn't specify what for. I understand that this is the top 10 chemicals of most concerns, but that should be clarified or removed. I would recommend removing this as the first sub-sentence makes the first sentence very heavy.

Author response

Thank you for this helpful suggestion. We agree that the original opening sentence was overly heavy and that the WHO reference was not sufficiently specific in this context. In response, we removed this phrasing and revised the sentence for improved clarity and readability.

Revision made

Mercury (Hg) occurs naturally in both inorganic and organic forms in the environment.

- Line 33: Pre 1450 is not pre anthropogenic but preindustrial.

Author response

Thank you for this helpful clarification. We agree and have revised the terminology accordingly, replacing “pre-anthropogenic” with “pre-industrial” in the manuscript.

Revision made

Accounting for both historical anthropogenic inputs and their continued re-emission, current atmospheric Hg concentrations are estimated to be approximately 450 % higher than pre-1450 (pre-industrial) levels, accompanied by an average 300 % increase in global Hg deposition and a 230 % increase in surface ocean Hg concentrations (Outridge et al. 2018). This enrichment is spatially uneven, with atmospheric Hg levels increasing by up to 1600 % in the Northern Hemisphere and up to 400 % in the Southern Hemisphere relative to pre-industrial levels (Li et al. 2020).

- Line 44 states once deposited but not specific about where. This makes the sediment in the next sentence sound like sediment in general but based on the rest of the sentence it sounds like the author specifically refers to sediment in water. Hg would not methylate for example in Saharan sediment.

Author response

Thank you for this helpful comment. We agree that the phrase “once deposited” was vague and could imply deposition into sediments in general, rather than specifically into aquatic environments where Hg methylation commonly occurs. We have revised the text to clarify that the conversion of inorganic Hg to methylmercury primarily occurs after deposition into aquatic systems, including aquatic sediments and the water column.

Revision made

Once deposited into aquatic ecosystems, inorganic Hg can be converted into methylmercury (MeHg), a potent neurotoxin (Gilmour et al. 2013). This process, facilitated by microbial activity, occurs in anoxic-suboxic sediments (Dai et al. 2021; Chen et al. 2025; Zhou et al. 2025) and within the water column of lakes....

- Line 47: Biomagnification is a sub process of bioaccumulation, so it doesn't bioaccumulate and biomagnifies. It bioaccumulates through bioconcentration (bioaccumulation through direct uptake) and biomagnification (bioaccumulation through trophic interactions)

Author response

Thank you for this helpful comment. We agree that the original wording was imprecise, as biomagnification is a trophic mechanism contributing to overall bioaccumulation rather than a separate parallel process. We have revised the sentence to more accurately describe how MeHg accumulates in aquatic organisms through direct uptake and trophic transfer across food webs..

Revision made

As MeHg enters aquatic food webs, it bioaccumulates through both bioconcentration (direct uptake) and biomagnification (trophic interactions), posing risks to fish and humans consuming fish (Al-Sulaiti et al. 2022; Basu et al. 2023).

- Line 123: you stated the analysis was performed using Python but is insufficient to reproduce it. The package or method used should be specified and ideally both the versions of both the packages and Python that was used for the analysis. Like you do in line 130 for QGIS I would consider it much easier to reproduce.

Author response

Thank you for this helpful suggestion. We agree that the original description was not sufficiently detailed to support reproducibility. We have revised the Methods section to specify the Python environment and the principal packages used for the PCA and associated statistical analyses, in a manner consistent with the software detail provided for QGIS.

Revision made

To explore multivariate relationships among predictor variables, Principal Component Analysis (PCA) was conducted on standardized data using Python 3.12.13 within the Google Colab environment, employing the scikit-learn, pandas, NumPy, SciPy, and matplotlib packages..

- The formatting of Tables 1–4 does not work in the manuscript's current layout. When printed, the table headers appear on one page while the corresponding values appear on the next, making the tables difficult to read. For example, in Table 1 the column header "reference" is split across lines, and several header rows span up to seven lines. In addition, the citation text within the tables often occupies four to five lines because the table columns are too narrow. I recommend either moving these tables to the Supplement, reformatting them in landscape orientation, or reducing the amount of information included so that the tables become readable.

Author response

Thank you for this helpful comment. We agree that the original formatting of Tables 1–4 was difficult to read within the manuscript layout, particularly due to narrow column widths, line-wrapped headers, and citation-heavy entries. In response, we have substantially revised the presentation of these tables to improve readability and reduce visual clutter.

Former Tables 1–3 were reorganized and consolidated into two revised tables with a simplified structure and improved formatting, while former Table 4 has been moved to the Supplementary Material. These changes improve readability in both on-screen and printed formats.

Revision made

Table 1. Projected changes in global anthropogenic mercury emissions and atmospheric deposition by 2050 under different policy scenarios and modeling frameworks.

Base year	Global anthropogenic Hg emissions in base year (Mg yr^{-1})	Global anthropogenic Hg emissions in 2050 (Mg yr^{-1})	Change in total Hg deposition in 2050 w.r.t. base year (%)	Model used for deposition simulations	Regions analyzed	References

SRES: Special Report on Emissions Scenarios—an Intergovernmental Panel on Climate Change (IPCC, 2000) framework that projects future emissions based on trajectories of energy use, fuel consumption, economic development, and technological change. The scenarios are grouped into four main families: A1, A2, B1, and B2.						
A1B: A subset of A1 where all energy sources are balanced (neither high fossil dominance nor full non-fossil). [A1: A future characterized by rapid economic and technological growth, with global population peaking mid-century]						
2000	2190	4856	75 75	CAM-Chem/Hg	Eastern U.S. Western U.S.	(Lei et al. 2014)
2005	1900	4300	21 30	GEOS-Chem	Global US	(Corbitt et al. 2011)
2010	1890	4900	87	GEOS-Chem + MITgcm + GTMM	Global	(Y. Zhang et al. 2021)
2015	2500	4900	40	5-box geochemical model for Arctic + Global Box Model	Arctic	(Chen et al. 2018; AMAP/UNEP, 2018)
2005	1900	4300	37	GEOS-Chem	Great Lakes	(H. Zhang et al. 2021)
A2: A fragmented, self-reliant world with high population growth, slow technological change, and regionally oriented development.						
2005	1900	3400	25	GEOS-Chem	Global	(Corbitt et al. 2011)

2010	1890	3900	59	GEOS-Chem + MITgcm + GTMM	Global	(Y. Zhang et al. 2021)
2006	2480	3900	12	CTM-Hg, TEAM, AERMOD	Mendums Pond, NH	(Vijayaraghavan et al. 2014)
A1F1: subset of A1 ("Fossil-intensive") where future energy use remains heavily dependent on fossil fuels.						
2000	2190	5984	100 75	CAM-Chem/Hg	Eastern U.S. Western U.S.	(Lei et al. 2014)
No policy: Assumes no new Hg or air-quality controls beyond those implemented by 2010, with continued coal combustion and limited emission-control expansion.						
2005-2006	2000	4140	30	GEOS-Chem	U.S.	(Giang and Selin 2016)
B1: An environmentally sustainable, convergent world with service-based economy and clean, efficient technologies.						
2000	2190	2386	13 0	CAM-Chem/Hg	Eastern U.S. Western U.S.	(Lei et al. 2014)
2005	1900	1900	1 -10 -22	GEOS-Chem	Global U.S. Northeast U.S.	(Corbitt et al. 2011)

2005	1900	1900	-13	GEOS-Chem	Great Lakes	(H. Zhang et al. 2021)
B2: A local sustainability-focused scenario with moderate growth and slower, diverse technological change.						
2005	1900	2200	7	GEOS-Chem	Global	(Corbitt et al. 2011)
2006	2480	2630	-15	CTM-Hg, TEAM, AERMOD	Mendums Pond, NH	(Vijayaraghavan et al. 2014)
Minamata: Global treaty scenario achieving major emission cuts via best available technologies and control measures across key Hg sources.						
2005-2006	2000	2270	5	GEOS-Chem	U.S.	(Giang and Selin 2016)
NPS: Implementation of pledged global actions (e.g., Minamata, fossil-fuel phase-outs) achieving notable emission cuts by 2035. [CPS: Continuation of 2010 policies and controls without new Hg-specific or climate initiatives.]						
2010	1960	0	-14 26 -55 -13 -15	GEOS-Chem + GBC Model	Global Ahmedabad (India) Shanghai (China) South Pacific Eastern lakes, Maine (U.S.)	(Angot et al. 2018)

2010	1890	1020	-28	GEOS-Chem + MITgcm + GTMM	Global	(Y. Zhang et al. 2021)
Constant emissions: Assumes no change in anthropogenic Hg emissions by 2050 with respect to 2015						
2015	2500	2500	12	5-box geochemical model for Arctic + Global Box Model	Arctic	(Chen et al. 2018; AMAP/UNEP, 2018)
MFR: Most optimistic case with universal adoption of best available technologies for maximum emission reduction.						
2010	1890	300	-48	GEOS-Chem + MITgcm + GTMM	Global	(Y. Zhang et al. 2021)
2010	2000	0	-24 -38 -68 -22 -26	GEOS-Chem + GBC Model	Global Ahmedabad (India) Shanghai (China) South Pacific Eastern lakes, Maine (U.S.)	(Angot et al. 2018)
Hg Controls: Assumes 50% reduction in primary anthropogenic Hg emissions by 2050 via widespread Hg-specific controls.						
2015	2500	1250	-16	5-box geochemical model for Arctic + Global Box Model	Arctic	(Chen et al. 2018; AMAP/UNEP, 2018)

Zero Emissions: Idealized case where all primary anthropogenic Hg emissions cease after 2015, representing the maximum achievable reduction.					
2015	2500	0	-50	5-box geochemical model Arctic for Arctic + Global Box Model	(Chen et al. 2018; AMAP/UNEP, 2018)

Table 2. Projected changes in atmospheric Hg deposition and corresponding fish MeHg concentrations by 2050 across different scenarios and modeling frameworks.

Base year	Change in total Hg deposition in 2050 w.r.t. base year (%)	Model used for simulations	Change in MeHg in fish in 2050 w.r.t. base year (%)	Models used for MeHg in fish simulations	Regions analyzed	Fish type	References
SRES: Special Report on Emissions Scenarios—an Intergovernmental Panel on Climate Change (IPCC, 2000) framework that projects future emissions based on trajectories of energy use, fuel consumption, economic development, and technological change. The scenarios are grouped into four main families: A1, A2, B1, and B2.							
SRES A2: A fragmented, self-reliant world with high population growth, slow technological change, and regionally oriented development.							
2006	12	CTM-Hg, TEAM, AERMOD	5	D-MCM	Mendums Pond, NH	LMB and yellow perch	(Vijayaraghavan et al. 2014)
SRES B2: A local sustainability-focused scenario with moderate growth and slower, diverse technological change.							

2006	-15	CTM-Hg, TEAM, AERMOD	-14	D-MCM	Mendums Pond, NH	LMB and yellow perch	(Vijayaraghavan et al. 2014)
NPS: Implementation of pledged global actions (e.g., Minamata, fossil-fuel phase-outs) achieving notable emission cuts by 2035. [CPS: Continuation of 2010 policies and controls without new Hg-specific or climate initiatives.]							
2010	-15	GEOS-Chem GBC Model	+-11	Hendrick's model	LMB	Eastern Maine lakes,	Brook trout, brown trout, burbot, Landlocked salmon, and smallmouth bass (Angot et al. 2018)
Policy-in-action: Full implementation of existing Hg-control measures, including Minamata Convention and U.S. Clean Air Act rules, stabilizing emissions near current levels by 2050.							
2005-2006	-15	GEOS-Chem	-11	Hendrick's model	LMB	Lakes in Michigan's UP	Northern pike, LMB, Yellow perch, and Pickerel (Perlinger et al. 2018)
	12		19			Lake in Michigan's UP with Adirondack's deposition	Walleye

	12		-38		Lake in Michigan's UP with Adirondack's deposition and watershed features	Walleye	
MFR: Most optimistic case with universal adoption of best available technologies for maximum emission reduction.							
2010	-26	GEOS-Chem GBC Model	+-22	Hendrick's model	LMB Eastern Maine lakes,	Brook trout, brown trout, burbot, Landlocked salmon, and smallmouth bass	(Angot et al. 2018)
Aspirational: Assumes complete elimination of anthropogenic Hg emissions by 2050 through a global transition to Hg-free technologies and renewable energy sources.							
2005-2006	-65	GEOS-Chem	-65	Hendrick's model	LMB Lakes in Michigan's UP	Northern pike, LMB, Yellow perch, and Pickerel	(Perlinger et al. 2018)
Reduction in deposition, 50%: A scenario applying a uniform 50% reduction to locally observed Hg deposition (without specifying global emissions).							
2001	-50	Community Multi-scale Air	-31		Eagle Butte Lake, SD	Northern pike	(Knightes et al. 2009)

Quality (CMAQ)	Model-51		Lake Barco, FL	LMB
	-57	WASP, SERAFM, BASS	Pawtuckaway Lake, NH	yellow perch
	-77		Lake Waccamaw, NC	LMB
	-74		Brier Creek, GA	Pickerel

- Line 204: The conclusion that changes in deposition are smaller than changes in emissions is expected. A significant portion of deposition originates from re-emissions and legacy Hg pools, so reductions in fresh anthropogenic emissions do *not* translate 1:1 into deposition changes.

Author response

Thank you for this insightful comment. We agree that the smaller response of Hg deposition relative to emissions is expected and does not reflect a one-to-one relationship. As noted by the reviewer, deposition is not solely controlled by current anthropogenic emissions but is also influenced by the global atmospheric Hg pool, including contributions from legacy reservoirs and re-emission processes. In response, We have revised the text to explicitly acknowledge this mechanism alongside the physical processes that influence the fate of emitted Hg.

Revision made

The relationship is strikingly linear ($R^2 = 0.79$, slope = 0.41, $p < 0.001$), meaning that 79 % of the observed variability in Hg deposition is explained by changes in emissions. Yet, the slope indicates that a 1 % change in emissions yields only about a 0.4 % change in deposition, implying that a substantial fraction of emitted Hg undergoes chemical transformation and/or transport, while also reflecting the influence of re-emissions and the legacy mercury pool.

- Lines 234 - 266: This section includes an interesting discussion, but is too long and geographically unfocused (China, Southern Hemisphere) compared to the rest of the paper which focusses on the US. Recommend shortening and focusing on the core point: observed nonlinearity vs. modeled linearity.

Author response

Thank you for this thoughtful comment. We believe the intended line range may be Lines 244–266 rather than 234–266, as this more directly corresponds to the discussion of international observational examples referenced in the comment. We agree that this section should remain focused on the manuscript's central point regarding the contrast between observed complexity and modeled linearity. In response, we have revised this section to improve focus and reduce its length, as suggested. We streamlined the discussion to emphasize the core point while removing less relevant detail. In addition, we incorporated additional examples from North America, particularly

the United States, to provide a more balanced geographical context alongside the previously included international cases (e.g., China).

These revisions aim to retain the broader relevance of the comparison while ensuring that the discussion remains clearly aligned with the main objectives of the manuscript.

Revision made

Model–observation inconsistencies are evident across multiple spatial scales. (Holloway et al. 2012) found that across 31 Mercury Deposition Network (MDN) sites in the Great Lakes region, the CMAQ-Hg model underestimated annual Hg wet deposition by about 21 % on average, with seasonal biases including underprediction in spring–fall and overprediction in winter (approx. 70 %). Similarly, MDN observations in the southeastern U.S. show that the global GEOS-Chem model (4°×5°) underestimated deposition by about 46 %, largely due to its inability to resolve deep convection, a key driver of Hg wet deposition (Xu et al., 2022). In China, observations indicate faster declines than models predict, with GEM decreasing by approx. 19 % compared to an 11 % modeled reduction over 2013–2017 (Feng et al., 2024; Liu et al., 2019).

Model-observation discrepancies are also apparent at the hemispheric scale. (Travnikov et al. 2017) evaluated four global Hg models (GLEMOS, GEOS-Chem, GEM-MACH-Hg, and ECHMERIT) against observations and found that performance depends strongly on oxidation chemistry and precipitation scavenging. Br-based schemes peak in spring, while OH-based schemes better capture North American and European summer peaks, indicating no single pathway fully explains observed patterns. Furthermore, in the Southern Hemisphere, GEOS-Chem and MITgcm overestimate wet deposition by approx. 5-fold and underestimate dry deposition by 3.5-fold relative to observations, reflecting simplified representations of convective uplift and precipitation scavenging (Leiva González et al., 2022).

References:

Holloway, T., C. Voigt, J. Morton, S. N. Spak, A. P. Rutter, and J. J. Schauer. 2012. "An Assessment of Atmospheric Mercury in the Community Multiscale Air Quality (CMAQ) Model at an Urban Site and a Rural Site in the Great Lakes Region of North America." *Atmospheric Chemistry and Physics* 12 (15): 7117–33. https://doi.org/10.5194/acp-12-7117-2012.

Xu, Xiaotian, Xu Feng, Haipeng Lin, et al. 2022. "Modeling the High-Mercury Wet Deposition in the Southeastern US with WRF-GC-Hg v1.0." *Geoscientific Model Development* 15 (9): 3845–59. https://doi.org/10.5194/gmd-15-3845-2022.

- Line 304: Growth dilution does not eliminate MeHg, it only reduces concentration.

Author response

Thank you for this helpful comment. We agree that growth dilution does not represent true Hg elimination, but rather a reduction in tissue concentration resulting from increases in body mass. We have revised the sentence to distinguish Hg elimination processes from growth-related dilution effects.

Revision made

Hg levels are moderated by metabolic elimination, which removes Hg, and growth dilution, which reduces MeHg concentrations (Barber 2008).

- Line 328: Interpretation of R^2 is incorrect. Since the axes represent *percent change in deposition* vs. *percent change in fish MeHg*, $R^2 = 0.63$ means 63% of the variance in the *change* of fish MeHg bioaccumulation is explained by the *change* in Hg deposition. Not 63% of the MeHg variance in fish. This distinction is important because each point in the plot represents a difference between two scenario simulations, meaning the underlying variance in fish MeHg itself is not present in the dataset.

Author response

Thank you for pointing out this important distinction. We agree that because our analysis uses percent changes between scenarios, the R^2 value represents the explained variance in the response (or change) of fish MeHg, rather than the variance of absolute MeHg concentrations. We have revised the text to explicitly state that.

Revision made

Despite the diversity of atmospheric models, bioaccumulation frameworks, fish species, and ecosystem types represented, the relationship is strongly linear ($R^2 = 0.63$, slope = 0.87, $p < 0.01$), indicating that roughly 63 % of the variance in the projected change of fish MeHg is explained by scenario-based changes in atmospheric Hg deposition.

- Section 3.2: Model a) and b) are marked with: while model d) and c) are marked with –. This should be uniform.

Author response

Thank you for this helpful observation. We agree that the notation used to distinguish model descriptions in Section 3.2 was inconsistent. We have revised the formatting so that all model descriptions are now introduced consistently using ":".

- Figure 3 and line 200: It should be stated explicitly how the percentage differences were calculated, including the exact equations used. More importantly, the manuscript must clearly clarify between which scenarios the differences were computed between. I initially assumed that each point represented a comparison within the same model under different emission scenarios, but this cannot explain the cluster of points where Hg emissions show no change while deposition increases or decreases. This suggests that the comparisons may instead be made against a baseline model or a model average and some of the differences in model output are caused by other drivers than Hg emissions, but I did not clearly understand this from the text. If so, the procedure and its implications should be clearly described, as the choice of reference scenario can strongly influence the resulting patterns.

Author response

Thank you for this important and helpful comment. We agree that the original manuscript did not describe the calculation of percentage differences with sufficient clarity, particularly regarding the equations used and the reference scenarios against which changes were computed. In response, we have added the corresponding equations and calculation details in the Supplement to improve transparency while keeping the main text concise.

Specifically, for each compiled scenario, percentage change was calculated relative to the corresponding base year used in the original study, rather than against a single harmonized baseline across all studies.

We also appreciate the reviewer's observation regarding the cluster of points where Hg emissions show little or no change while Hg deposition either increases or decreases. These points do not

necessarily represent paired comparisons within the same model under otherwise identical conditions. Rather, they were compiled across different published studies, which may differ in model structure, emissions inventories, oxidation chemistry, spatial resolution, meteorological forcing, and baseline assumptions, even when nominally evaluating similar policy scenarios, years, or regions. As a result, similar or even identical percentage changes in emissions can still produce different modeled deposition responses across studies. This source of divergence was already explained in the original manuscript (Lines 222–227), where we noted that even under relatively stable emission scenarios, modeled deposition responses can differ across studies due to differences in model setup, spatial resolution, inventories, and assumptions.

We anticipate that the addition of the explicit equations and calculation details in the Supplement will now make this framework clearer.

Revision made

We revised the Methods section to explicitly define how percentage changes were calculated and to clarify that each point represents a scenario-based comparison relative to the base year used in the original source study.

Percentage annual change for each variable (X = Hg emissions, Hg deposition, or fish MeHg) was calculated as:

$$\%\ annual\ change\ in\ X\ = \left(\left(\frac{X_{2050,\,scenario}}{X_{base\ \mathrm{year},\,scenario}}\right)^{\frac{1}{\Delta t}} - 1\right) * 100$$

Where:

$X_{2050,\,scenario}\ =\ modeled\ value\ under\ a\ given\ scenario\ in\ 2050$

$X_{base\ \mathrm{year},\,scenario}\ =\ modeled\ value\ in\ the\ base\ year\ used\ in\ that\ study$

$\Delta t\ =\ number\ of\ years\ between\ the\ base\ year\ and\ future\ year$

- Figure 4: In the scenarios where Hg deposition is expected to increase, there are two points labeled *L-HTL b+3* that share the same color but show very different changes in fish MeHg bioaccumulation. The manuscript should clarify why two identically labeled scenarios produce such different outcomes. In addition, the colors are described as

representing deposition-reduction categories, but the x-axis already represents the change in deposition. This creates redundancy and confusion. Finally, several points on the far right of the plot share the same change in Hg deposition but are assigned to different reduction categories (e.g., "middle" vs. "high"). It is unclear how the same deposition change can belong to multiple deposition-reduction classes, and the criteria for assigning these categories should be explained.

Author response

Thank you for this careful and insightful comment. We agree that several aspects of Figure 4 were not sufficiently clear and could lead to confusion.

Regarding the "L-HTL b+3" points, we acknowledge that the labeling and presentation were unclear. These points represent results from the same Policy-in-Action scenario but under different watershed configurations. As shown in Perlinger et al. (2018), even under identical atmospheric deposition conditions, variations in watershed characteristics (e.g., wetland extent and runoff pathways) can lead to substantially different fish MeHg responses.

Regarding points on the far right of the plot that share similar deposition changes but fall into different reduction categories (e.g., "middle" vs. "high"), We clarify that these categories were not assigned by us based on a uniform numerical threshold in Hg deposition change. Rather, they were adopted from the original scenario classifications used in the source studies, where scenarios were defined according to each study's own policy or emissions framework, including SRES, Policy-in-Action, MFR, NPS. For example, the SRES A2 scenario, represented by the pink circle, is a high-emission scenario designed to project increasing emissions and deposition, whereas the Policy-in-Action scenario represents a moderate deposition-reduction pathway. In this specific case, changing the regional setting caused the modeled Hg deposition under the Policy-in-Action scenario to increase to a level comparable to that projected under the increasing-deposition SRES A2 scenario, despite their different original scenario classifications.

Likewise, under the same Policy-in-Action deposition scenario, the modeled decline in fish MeHg for lakes in Michigan's Upper Peninsula is moderate (-0.25 % yr^{-1}). However, in this specific case, altering the regional and watershed characteristics shifted the modeled fish MeHg response in opposite directions: in one case, from a decrease to an increase, and in the other, to a stronger decrease. This highlights the strong influence of watershed sensitivity on fish MeHg response, even under the same deposition scenario. As a result, in Figure 4, similar Hg deposition outcomes

may correspond to different scenario categories, while contrasting fish MeHg responses may still occur under the same policy scenario when watershed characteristics (e.g., wetland percentage and forest runoff) differ.

We have clarified this distinction in both the figure caption and the main text. In response to these comments, we revised Figure 4 and the corresponding text (Lines 334–337) to more clearly distinguish between scenario-driven deposition changes and watershed-driven variability.

Revision made

For example, under the intermediate Policy-in-Action scenario with moderate deposition reductions, the modeled decline in fish MeHg for lakes in Michigan's Upper Peninsula (UP) is moderate. However, two virtual experiments reported by Perlinger et al. (2018) separate the effects of scenario-based atmospheric deposition and watershed sensitivity on fish MeHg responses (Ellipse A; symbol ∇). Applying Adirondack deposition to the Michigan UP lake resulted in increased modeled fish MeHg response, whereas applying Adirondack watershed characteristics under the same deposition conditions led to a decrease, highlighting the influence of watershed properties. Similarly, under the high deposition-reduction scenario involving a 50 % decrease in Hg deposition (Ellipse B; symbol Δ), five integrated data points show substantial variability in fish MeHg responses under the same policy scenario (Knightes et al., 2009).

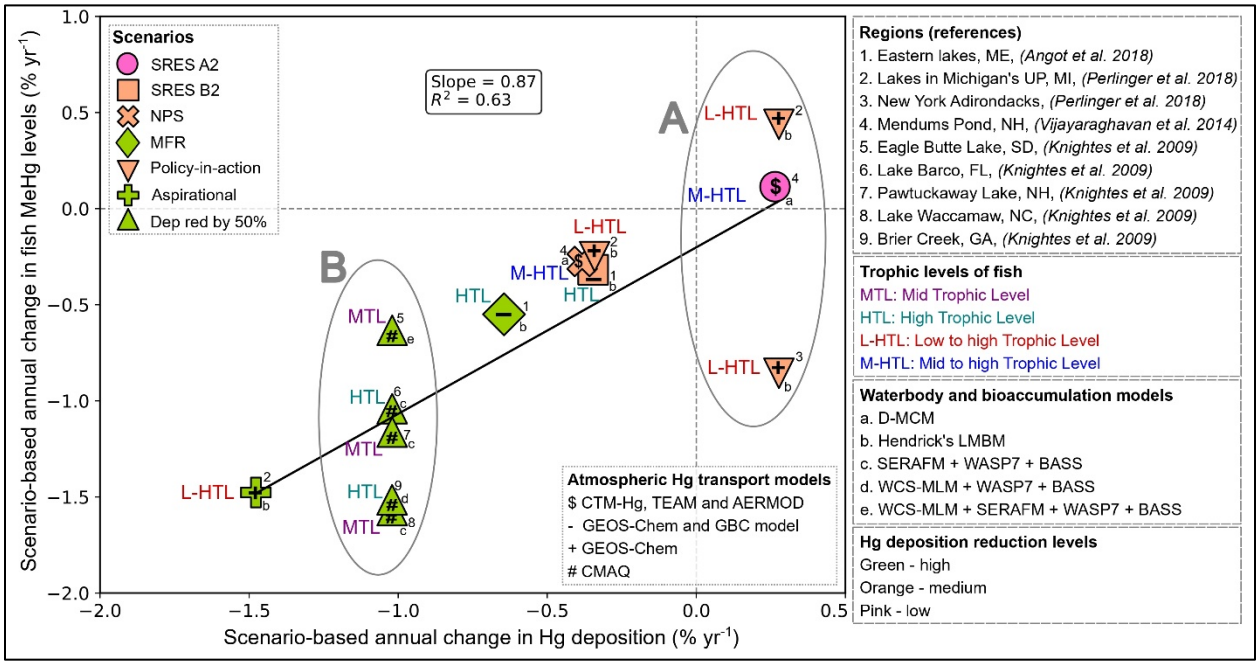

Figure 4: Relationship between scenario-based projected annual changes in atmospheric Hg deposition (% yr^{-1}) and corresponding changes in fish MeHg concentrations (% yr^{-1}) by 2050. Marker shapes (top left panel) denote

Hg deposition scenarios, while special characters (bottom right panel) represent atmospheric Hg deposition models. Numbers shown in the top right panel correspond to study regions and references. Fish trophic levels are color-coded in the lower panel, and the bottom right panel also indicates waterbody and watershed type together with the bioaccumulation models used. Marker fill colors indicate the magnitude of Hg deposition reduction, classified as high (green), medium (orange), and low (pink).

- Figure 5: I believe this figure should be removed or substantially revised. Currently, the figure is confusing and did not become clear until reading the text *after* it. That text makes the important point that models often predict a limited, linear response while bioaccumulation in nature can follow very different patterns. The text used excellent references from relevant geographical locations, but in my opinion the figure itself does not help illustrate that conclusion. It was also initially confusing that models are shown by a bar, but observational data is 2 points and 1 bar with error bars. Only after reading the text and comparing the years did I understand what meant what, from the figure itself this was not clear, and after reading the text the figure did not improve my understanding.

Author response

Thank you for this helpful comment. We agree that Figure 5 was not sufficiently intuitive and that its interpretation relied too heavily on the surrounding text. We also agree that, in its current form, the figure did not effectively strengthen the main conclusion regarding the contrast between the relatively linear responses projected by current models and the more variable patterns observed in empirical studies. Following the reviewer's suggestion, we have removed Figure 5 from the manuscript. The key observational data and references previously illustrated in the figure were already included in the main text and remain unchanged.

Revision made

Figure 5 has been removed.

- The title of Section 3.3, *"Lake characteristics vs MeHg in fish, 2050,"* could be improved for clarity and readability. I recommend rephrasing it to more clearly reflect

the content, for example to “Influence of Lake Characteristics on Projected 2050 Fish MeHg Levels”.

Author response

Thank you for this helpful suggestion. We agree that the original section title could be improved for clarity and readability. In response, we revised the title of Section 3.3.

Revision made

The section titles have been revised as follows:

- 3.3 Lake characteristics vs MeHg in fish, 2050 → Influence of lake characteristics on projected fish MeHg levels.

Structural comments

- I recommend shortening the conclusion and adding a separate Discussion section, or a section purely focused on lesson for future lake Hg modeling. While the manuscript correctly notes that the models used underrepresent natural complexity, it does not provide a mechanistic explanation of why this occurs or in-depth discussion of why this is, and what it means. For example, the Hendricks and SERAFM models simulate bioaccumulation using BAF-based relationships, which are inherently linear; therefore, unless I misunderstand, any non-linearity between Hg deposition and fish MeHg must arise from Hg speciation processes rather than from the bioaccumulation step itself as the BAF is a linear relationship between marine MeHg and fish MeHg, as per definition. In models you should always make sure not to rediscover your own parameterization, and observing linear behavior in a BAF based models should not be surprising. More broadly, all ecosystem models included in this analysis are box models, which cannot represent several hydrodynamic processes known to influence both MeHg bioaccumulation and ecosystems interactions. I think this raises 2 key questions that should be discussed. The first is the question if the MeHg models are actually wrong. If a lake model demonstrates that a 2% reduction in deposition in the lake would result in a 2% reduction in fish MeHg and we assume for a moment the model is completely correct. Then a change

in trophic structure in the lake can still cause the observed difference in fish MeHg to be different than predicted. But this does not mean the MeHg bioaccumulation model is per se wrong. It means that MeHg bioaccumulation in lakes can be improved using coupled mechanistic models, rather than box models.

A key question that follows is how these models can be improved. The dataset already contains useful information that could guide such a discussion. For instance, the PCA shows that water depth is negatively correlated with the percent change in fish MeHg, while watershed area is positively correlated. This is intuitive: the ratio of atmospheric deposition area to total water volume likely influences a lake's sensitivity to changes in deposition. Exploring these relationships more fully, with a direct focus on how to contribute to the observed nonlinearity and what this means for models would add depth to the analysis and help translate the findings into concrete recommendations for improving future mercury bioaccumulation models. This is somewhat discussed in section 3.3, but here the focus is mostly on statistical analysis of what drives MeHg concentrations in fish in lakes, and I think a clearer direct coupling to what lessons can be learned from the collected data can greatly improve this manuscript.

I understand this may not be possible and I would support the publishing without this. But it is noteworthy that the link between Hg emission and deposition in most models are based on complex 3D atmospheric models that are in general well understood and validated. Whereas the lake component of this analysis links Hg deposition to MeHg in fish using much simpler models, even though this pathway is poorly understood and extremely complicated. If the data is available splitting the lake component into Hg deposition to marine MeHg concentrations and marine MeHg concentrations to MeHg in fish the study could help to identify which components are nonlinear in nature and linear in the models, and thus what components of the models need to be improved.

Author response

Thank you for this thoughtful and highly constructive comment. We appreciate the reviewer's detailed insights regarding the mechanistic interpretation of model behavior and the implications for future Hg bioaccumulation modeling.

We agree that clarifying the basis of the observed linear responses improves the manuscript. At the same time, we note that the relatively linear behavior is consistent with the structure of many existing models, including steady-state or BAF-based formulations, which impose proportional relationships between aqueous MeHg and fish MeHg. As such, these outcomes are not unexpected and reflect the intended level of simplification rather than a flaw in the models themselves.

In response to the reviewer's suggestion, we have added a brief, focused paragraph in Section 3.3 (Pages 26–27) to clarify how model assumptions—such as simplified watershed transfer, limited representation of legacy storage, and reduced ecological complexity—may contribute to differences between modeled and observed responses. This addition also links these concepts to the PCA results (e.g., watershed area, wetland extent, and lake morphometry) to provide context for the observed sensitivities.

While we agree that further development of fully coupled mechanistic frameworks is an important direction, we have intentionally limited this discussion to a concise interpretation rather than proposing detailed model improvements, as the aim of this synthesis is to evaluate patterns across existing studies rather than prescribe specific model formulations.

Regarding the reviewer's suggestion to further decompose the pathway into deposition → aqueous MeHg and aqueous MeHg → fish MeHg components, we agree that this is an important direction for future work. However, such separation was not feasible within the scope of this synthesis due to limited and inconsistent reporting of intermediate variables across the compiled studies.

Overall, we have aimed to incorporate these perspectives in a focused manner while maintaining the scope of the synthesis, and we are grateful for the reviewer's suggestions, which have helped strengthen the interpretation and clarity of the manuscript.

Revision made

The negative association between wetland extent and projected declines in fish MeHg observed in the PCA contrasts with many field studies that report elevated absolute fish MeHg with increasing wetland coverage, although the magnitude and direction of this relationship vary with sulfate availability, dissolved organic matter, and hydrologic setting (Watras et al. 2005; Ackerman et al. 2019; Poulin et al. 2025). This apparent discrepancy likely reflects model assumptions regarding Hg loading pathways and response kinetics. Modeling frameworks derived from or conceptually similar to SERAFM assume relatively direct transfer of atmospherically deposited Hg from the catchment to the lake, with limited representation of hydrologic retention

and delayed release from soils and wetlands (Knightes et al., 2008). Under this structure, wetlands and large watersheds act as highly sensitive receptors of atmospheric Hg inputs, strengthening atmospheric–catchment coupling in wetland-dominated systems. Our supplemental regression analysis supports this interpretation, showing that watershed and wetland area are strong predictors of total Hg loading to the lake ($R^2 = 0.90$ and 0.87, respectively; $p < 0.0001$). In addition, PCA results indicate that lake morphometry controls model sensitivity, with large, shallow lakes responding strongly to deposition changes due to limited dilution. Fish MeHg concentrations are commonly estimated using a steady-state BAF approach, which does not capture the biological and ecological lag times required for fish populations to fully adjust to changes in water-column Hg exposure (Angot et al., 2018). Consequently, declines in modeled water-column Hg are translated into relatively rapid declines in fish MeHg, whereas in natural systems, Hg stored in wetland soils and catchment reservoirs can delay responses to emission reductions. Thus, although wetland-dominated systems often support higher absolute MeHg concentrations in empirical studies, they emerge in these simulations as the most dynamic responders to declining atmospheric inputs because legacy storage and delayed ecosystem adjustment are only partially represented. In contrast, emerging 3-D marine Hg models such as MERCY and OGSTM–BFM–Hg demonstrate how embedding Hg cycling and bioaccumulation in high-resolution physical–biogeochemical frameworks with multi-trophic food webs can resolve climate impacts and food-web–mediated exposure with far greater realism (Rosati et al. 2022; Bieser et al. 2023). Biogeochemical reaction rates and lake-chemistry variables (e.g., DOC, alkalinity, sulfur) could not be evaluated here due to data limitations but incorporating them in future work will be essential for linking model behavior with real-world ecosystem responses.

References

Bieser, Johannes, David J. Amptmeijer, Ute Daewel, Joachim Kuss, Anne L. Soerensen, and Corinna Schrum. 2023. "The 3D Biogeochemical Marine Mercury Cycling Model MERCY v2.0 – Linking Atmospheric Hg to Methylmercury in Fish." *Geoscientific Model Development* 16 (9): 2649–88. https://doi.org/10.5194/gmd-16-2649-2023.

Rosati, Ginevra, Donata Canu, Paolo Lazzari, and Cosimo Solidoro. 2022. "Assessing the Spatial and Temporal Variability of Methylmercury Biogeochemistry and Bioaccumulation in the Mediterranean Sea with a Coupled 3D Model." *Biogeosciences* 19 (15): 3663–82. https://doi.org/10.5194/bg-19-3663-2022.

Conclusion:

In summary, despite substantial differences among atmospheric models in resolution, emissions, meteorology, and redox chemistry, and among bioaccumulation and ecosystem representations, our synthesis shows they consistently predict near-linear relationships between emissions, deposition, and fish MeHg responses. Yet comparisons with observations reveal important departures from these idealized trends along both links of the causal chain (emissions → deposition and deposition → fish MeHg). In particular, Hg bioaccumulation models generally capture the direction of change under emission controls but underrepresent the ecological, biogeochemical, and regional heterogeneity that governs response rate and magnitude. Empirical evidence shows fish MeHg can respond nonlinearly, with lags or partial decoupling from declining deposition due to food-web restructuring, invasive species, wetland-mediated methylation, internal cycling, and climate-driven changes in productivity and hydrology. These findings highlight the need to pair global modeling with long-term atmospheric, depositional, and ecological observations, and to improve representation of watershed processes, food-web dynamics, and climate feedbacks, to more reliably quantify the benefits of emission reductions under the Minamata Convention.

Thus, although current models are powerful and indispensable tools, their orderly linearity invites reflection: do we truly capture the intricate feedbacks of the real atmosphere and natural ecosystems? To strengthen future assessments and policy strategies, we should enhance model complexity by incorporating nonlinear processes and bridge gaps with observations through prioritized field studies and long-term monitoring. By doing so, we can better validate projections, uncover unaccounted-for interactions, and ensure models evolve in step with the complexity of Hg cycling.

- A final point regards over-citation throughout the manuscript. It is of course good to cite previous literature, but I strongly recommend making a clear statement and supporting it with one or two relevant references, ensuring that it is evident which reference supports which claim. The most notable example is the citation block around line 73, which spans

four lines and includes roughly 20 papers, yet the rationale for including each citation and their specific contributions is unclear. This block refers broadly to previous modelling work, but the selection appears inconsistent: for instance, studies such as Zhang et al. (2020) are included even though they focus on marine 3-D modelling, while other relevant marine studies (e.g., Rosati et al. 2023; Bieser et al. 2023) are omitted. The problem is that if a reader wants to verify or research further in a citation block of 20 citations it is hard to know what the most relevant citations are. I suggest rewriting this section to clearly describe the types of models that exist, what they capture, and what is missing, while keeping citations closely tied to the specific statements they support. I would suggest this section should focus on lake and freshwater models, with marine models potentially introduced in the discussion, where they can illustrate processes or mechanistic complexity missing from current lake models. Similar over-citation occurs in other parts of the manuscript (often blocks of 5–10 references), which reduces readability and makes it difficult to evaluate the specific statements.

Author response

Thank you for this important and constructive comment. We agree that the previous version of the manuscript contained overly dense citation blocks, which could reduce clarity and make it difficult to identify the specific contribution of individual studies.

In response, we revised this section (previously around Line 73) to improve clarity and structure, particularly in relation to Hg modeling frameworks and freshwater/lake bioaccumulation models.

Rather than providing detailed descriptions of individual models already discussed later in the manuscript (Lines 275–322), we now provide a concise overview of model types, their key features, and their limitations. This revision aims to improve readability while more clearly conveying how different modeling approaches contribute to understanding Hg dynamics and where important gaps remain.

We have also reduced the number of citations and ensured that references are more directly linked to specific statements, highlighting the role and limitations of each model type. Marine modeling studies (e.g., Rosati et al., 2023; Bieser et al., 2023) have been incorporated into the Discussion

section (Section 3.3), where they are more directly relevant for comparison with broader ecosystem processes.

Revision made

Estimating future Hg emissions and their impacts on ecosystem and fish Hg levels requires consideration of evolving socio-economic and technological drivers, atmospheric chemistry, biogeochemical processes, climate variability, land-use change, and ongoing emissions and re-emissions from legacy sources. This complexity has led to diverse modeling approaches, from simple to integrated frameworks with varying process representation. For example, D-MCM models link atmospheric deposition, in-lake cycling, and bioenergetic food webs, but simplify watershed inputs and use fixed trophic groups (Vijayaraghavan et al., 2014). Similarly, BASS-based coupled models simulate loading and bioaccumulation, yet retain simplified watershed export and food-web dynamics with limited treatment of legacy Hg, DOC, and ecological variability (Knightes et al. 2009). More recent non-steady-state models, such as SERAFM-derived frameworks, incorporate hydrology and water–sediment interactions but still estimate fish Hg using bioaccumulation factors, limiting representation of trophic dynamics and lags. Notably, most lake models link watershed export directly to atmospheric deposition, without explicit representation of soil Hg storage or legacy release (Perlinger et al., 2018).

Overall, I enjoyed reading the manuscript and I think the core argument, that MeHg bioaccumulation models overpredict linearity compared to nature is an important message. However, I do believe the manuscript would be greatly improved if it translated these observations into a meaningful discussion about the quality of the models and what should be improved upon.

Author response

Thank you for this thoughtful and encouraging overall assessment of our manuscript. We appreciate the reviewer's recognition of the importance of the core finding regarding the tendency of current MeHg bioaccumulation models to produce relatively linear responses compared to observations.

In response to this comment, and in line with the reviewer's earlier suggestions, we have revised Section 3.3 to strengthen the mechanistic interpretation and broader implications of our results. Specifically, we expanded the discussion to explicitly address:

- the structural reasons why many current Hg bioaccumulation models produce relatively linear responses
- the key limitations of existing modeling frameworks, including simplified representations of trophic dynamics, watershed processes, and ecosystem interactions, and
- the implications for future model developments.

These revisions aim to more clearly translate the synthesis results into actionable insights for improving future Hg bioaccumulation modeling.