www.nature.com/scientificdata

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Global mercury dataset with predicted methylmercury concentrations in seafoods during 1995–2022

Haifeng Zhou[1], Yumeng Li[1,2], Qiumeng Zhong[2], Xiaohui Wu[1] & Sai Liang[2]✉

Mercury exposure poses significant threats to human health, particularly in its organic form, methylmercury (MeHg). Diet is the main pathway for human MeHg exposure, especially through seafood consumption. In this context, numerous studies have established seafood MeHg concentration datasets to assess MeHg-related health risks from seafood consumption. However, existing datasets are limited to specific regions and short-term observations, making it difficult to support continuous and dynamic assessments of global MeHg-related health risks. This study takes a bottom-up approach to construct a global seafood MeHg concentration dataset during 1995-2022. Firstly, it compiles a long-term time series marine-scale dataset of seafood MeHg concentrations, based on the reported seafood mercury concentrations from existing literature and machine learning methods. Subsequently, this study used the seafood catch volumes of each nation in different marine areas as weights to estimate the national-scale seafood MeHg concentrations. This dataset can provide essential data support for environmental impact assessment of mercury and its compounds as mentioned in Articles 12 and 19 of the Minamata Convention on Mercury.

Background & Summary

Methylmercury (MeHg), a highly toxic mercury (Hg) compound, can significantly impair the human nervous system[1], cardiovascular system[2,3], and immune system[4]. These health risks would further lead to economic losses. For example, Zhang *et al.* found that global economic loss induced by MeHg-related health risks could reach $117 billion annually[5]. In this context, over 150 nations and organizations have become signatories to the Minamata Convention on Mercury, which aims to mitigate the adverse impacts of Hg and its compounds on human health and the environment[6].

Human beings are exposed to MeHg through various pathways, including direct inhalation of anthropogenic Hg emissions and the use of Hg-containing products[3]. Diet is the primary route of human exposure to MeHg, particularly through seafood consumption[7]. For example, 87%-95% of total Hg exposure among U.S. adults is attributed to fish consumption[8]. Moreover, in coastal island nations (e.g., the Seychelles Islands), where seafood consumption is a staple in local diet, MeHg exposure levels among residents far exceed the general population[9]. Therefore, monitoring MeHg concentrations of seafoods is important. It helps assess MeHg-related health risks for populations and consequently provide scientific dietary guidelines for the public[10].

Over the past 30 years, numerous individual studies have monitored MeHg concentrations of seafoods. Meanwhile, researchers have compiled regional seafood MeHg concentration databases based on existing research data. However, these studies have primarily focused on economically developed regions (e.g., the U.S. and Europe) and high Hg emitters (e.g., China)[10–12]. Moreover, most studies have only reported MeHg concentrations of seafoods at a single time point or over a short time period. These limitations restrict the usefulness of MeHg concentration data in two aspects. On one hand, the geographical limitations prevent a comprehensive assessment of MeHg levels of seafoods worldwide, consequently hindering a systematic evaluation of global

[1]School of Environment, Beijing Normal University, Beijing, 100875, P. R. China. [2]Guangdong Basic Research Center of Excellence for Ecological Security and Green Development, Key Laboratory for City Cluster Environmental Safety and Green Development of the Ministry of Education, School of Ecology, Environment and Resources, Guangdong University of Technology, Guangzhou, 510006, P. R. China. ✉e-mail: liangsai@gdut.edu.cn

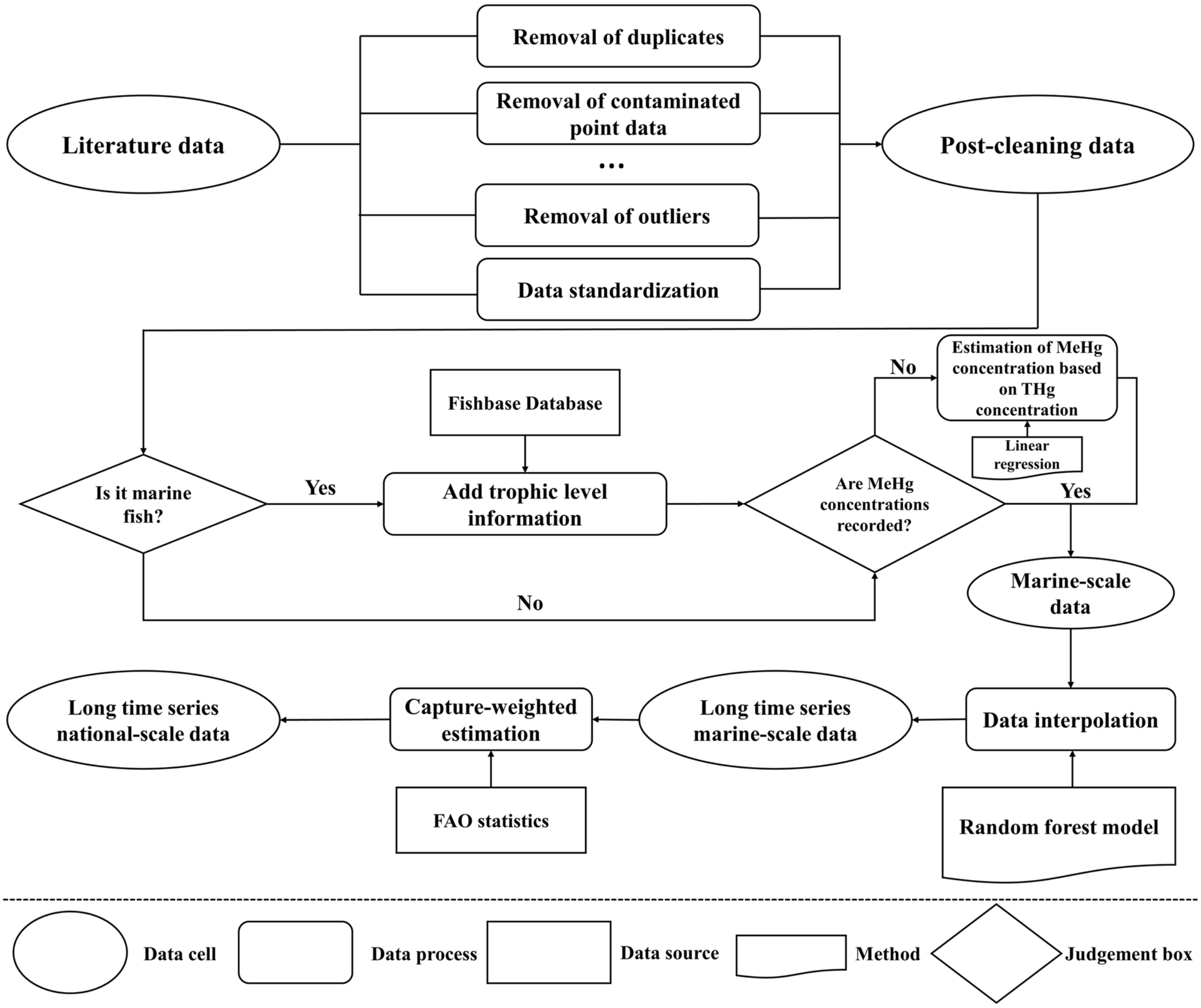

Fig. 1 Framework for constructing global seafood MeHg concentration dataset.

Contents	Data sources
Trophic levels of marine fishes	https://www.fishbase.se/search.php (available in January, 2025)
Seafood capture data	https://www.fao.org/fishery/en/collection/capture?lang=en (available in January, 2025)
FAO fishing areas	https://www.fao.org/fishery/en/area/search (available in January, 2025)

Table 1. Data sources of this study.

MeHg-related health risks. On the other hand, the limited time span of the data restricts the understanding of temporal trends in MeHg concentrations of seafoods. This makes it difficult to effectively identify and provide early warnings for high-risk regions dynamically.

To address the limitations of existing studies, this study constructed a global dataset of seafood MeHg concentrations during 1995-2022. For most nations, data on MeHg concentrations of their seafoods from existing research records are extremely limited. Thus, this dataset is constructed using a bottom-up approach. Specifically, we first compiled a dataset of seafood MeHg concentrations from various global marine areas during 1995-2022. Subsequently, the national MeHg concentrations of seafoods were estimated based on national fishing structure of seafoods (see details in Methods). This dataset provides an important data foundation for the assessment of environmental impacts of Hg and its compounds as pointed out in Articles 12 and 19 of the Minamata Convention on Mercury[6]. It can facilitate the development and adjustment of national food safety standards and public health policies.

Methods

The process of constructing global seafood MeHg concentration dataset is shown in Fig. 1. First, we constructed a long-term time series dataset of seafood MeHg concentrations from all marine areas based on existing literature and statistical methods. Subsequently, we estimated the national seafood MeHg concentrations based on fishing structure. The data sources of this study are shown in Table 1.

Compiling long-term time series marine-scale datasets of seafood MeHg concentrations. The construction of long-term time series datasets of seafood MeHg concentrations at the marine scale primarily involved the following steps: data collection, standardization, and classification; estimation of seafood MeHg concentrations at the marine scale; and imputation of missing data.

Collection of seafood MeHg concentrations. The total Hg (THg) and MeHg concentrations of seafoods were retrieved from peer-reviewed studies published between 1950 and 2024 in the Web of Science database. The literature search employed a Boolean combination of the following phrases: (1) "mercury" and (2) "fish OR seafood OR crustaceans OR mollusks". Subsequently, we applied the following criteria to screen the literature:

(1) Removing duplicate data records, such as duplicate original research data in review articles[10].
(2) Excluding data from sampling sites affected by point source pollution, because MeHg concentrations in these areas are typically higher and potentially lead to an overestimation of regional MeHg concentration in seafoods[5,10,13,14]. Only data points explicitly identified as originating from polluted areas were excluded, rather than entire studies. Polluted sites were defined based on descriptions in the source literature, such as proximity to industrial activities (e.g., in the influence range of gold mines, power plants, or chemical factories), evidence of environmental contamination (e.g., abnormal Hg concentrations in local biota), or locations in known high-risk zones (e.g., areas with a history of industrial discharge or mining activities). When such information was not clearly provided, data were retained to avoid unnecessary exclusions.
(3) Excluding MeHg concentrations from non-muscle tissues, because MeHg primarily accumulates in muscle tissue and data from other tissues may underestimate MeHg concentrations of seafoods[10].
(4) Removing data with unidentified marine areas.
(5) Removing data with unspecified weight type (i.e., dry weight or wet weight).
(6) Removing data below the detection limit.
(7) Removing 5% of outliers from the data using the interquartile range rule[15].

Standardization of seafood MeHg concentrations. To standardize Hg concentration data, we extracted and transformed the indicators described in Table S1 (see supplementary xlsx file). Specifically, we converted all concentration units to nanograms per gram (ng/g) and converted dry weight to wet weight, assuming a water content of 80%[10,13], as shown in Eq. 1[16]. For studies that only reported concentration means but not standard deviations, we assumed a standard deviation of 65% of the mean[14]. In addition, for studies that did not explicitly specify the sampling year, we assumed the actual sampling year to be two years prior to the publication year[17]. In cases where research only provides seafood MeHg concentrations for a specific time period, this study assumes that the seafood MeHg concentration in each year of that period is equal to the reported value. Finally, based on the sampling location information provided in the literature, we categorized the data according to the fishing areas defined by the Food and Agriculture Organization of the United Nations (FAO). The FAO fishing areas include Arctic Sea, Atlantic Antarctic, Atlantic Eastern Central, Atlantic Northeast, Atlantic Northwest, Atlantic Southeast, Atlantic Southwest, Atlantic Western Central, Indian Ocean Antarctic, Indian Ocean Eastern, Indian Ocean Western, Mediterranean and Black Sea, Pacific Antarctic, Pacific Eastern Central, Pacific Northeast, Pacific Northwest, Pacific Southeast, Pacific Southwest, and Pacific Western Central.

$$WW = (1 - 0.8) * DW \tag{1}$$

The notations WW and DW represent the wet weight and dry weight, respectively.

Classification of seafoods. There are significant differences in MeHg concentrations among different seafoods. Generally, MeHg concentrations of marine fishes are higher than those of other seafoods, particularly in predatory fishes[18,19]. Consequently, we classified seafoods into two groups: marine fishes and other seafoods. This classification facilitates a more detailed analysis and understanding of MeHg concentrations in different seafoods. Moreover, the vast diversity of seafood species hinders the construction of long-term MeHg concentration datasets for each individual species based on existing research. The trophic level is a crucial factor influencing seafood MeHg concentrations, with higher trophic levels generally exhibiting higher MeHg concentrations[20,21]. Following the approach by Zhang *et al.*[5], we classified marine fishes based on their trophic levels and constructed MeHg concentration datasets for different trophic level categories. Trophic levels of marine fishes were obtained from the Fishbase database (Table 1). Based on their trophic levels, we classified marine fishes into four categories: Level 1 (1.5-2.5), Level 2 (2.5-3.5), Level 3 (3.5-4.5), and Level 4 (4.5-5.5)[5]. For other seafoods, due to the lack of corresponding trophic levels in existing databases (e.g., Fishbase and Sealife databases), we categorized them into a single category of "seafoods (excluding marine fishes)".

Estimation of seafood MeHg concentrations at the marine scale. In most studies, only THg concentrations of seafoods were reported. To address the lack of direct MeHg data, we followed previous studies by converting THg to MeHg using a regression relationship derived from available data where both THg and MeHg values were reported[13,14] (Eqs. 2 and 3). To mitigate the uncertainties inherent in using predicted values rather than direct measurements, we employed a Monte Carlo simulation, running it 10,000 times to convert the THg data from studies that only reported THg values into corresponding MeHg estimates. The resulting mean and standard deviation from these simulations represent the uncertainty range of the MeHg data. Notably, most studies report the mean concentrations of all samples, rather than the concentrations of individual samples. Therefore, we recorded the sample size for each study and used these sample sizes as weights to estimate the mean concentrations of seafoods in each marine area (Eq. 4). For studies that did not provide sample size information, we assumed a sample size of 1 to reduce potential biases that such data might introduce when estimating the overall mean.

$$MeHg_{MFS} = e^{\ln(THg_{MFS}) - 0.23} \tag{2}$$

$$MeHg_{OTS} = e^{0.94*\ln(THg_{OTS}) - 0.47} \tag{3}$$

$$C_{ljt} = \frac{\sum_{i=1}^{n} C_{lijt} * m_{lijt}}{\sum_{i=1}^{n} m_{lijt}} \tag{4}$$

The notations $MeHg_{MFS}$ and $MeHg_{OTS}$ represent the MeHg concentrations of marine fishes and other seafoods, respectively, while THg_{MFS} and THg_{OTS} represent the THg concentrations of marine fishes and other seafoods, respectively; C_{ljt} represents the sample-weighted MeHg concentration of seafood l in marine area j at time t, while C_{lijt} represents the MeHg concentration of seafood l in marine area j in study i at time t; n represents the total number of studies; and m_{lijt} represents the sample size of seafood l in marine area j in study i at time t.

Imputation of missing concentration values. In existing studies, seafood MeHg concentrations are incomplete in certain years, and data prior to 1995 are particularly scarce (Fig. 2). Therefore, the long-term time series of seafood MeHg concentrations constructed in this study starts from 1995. We employed the random forest algorithm from machine learning techniques to impute the missing values. The random forest algorithm has been widely applied in the field of data interpolation[22–25]. Due to its outstanding non-linear processing capability, the random forest algorithm shows better performance in predicting complex time series data compared to traditional linear interpolation methods. In this study, we used the "RandomForestRegressor" module in the "sklearn.ensemble" package of Python 3.11 to implement this algorithm. We randomly divided the data into a training set and a test set with a ratio of 90% and 10%, respectively[22,23]. The training set was used for model construction and training, while the test set was used to evaluate the prediction accuracy of the model. The detailed implementation processes can refer to the source code in the Zenodo[26].

The robustness of the random forest simulation results will decrease when faced with severe data gaps. From the perspective of data availability, MeHg concentrations of Level 2 and Level 3 marine fishes are relatively complete (Fig. 2), allowing for direct interpolation using the random forest method. However, there is a significant lack of data on MeHg concentrations in Level 1 and Level 4 marine fishes, as well as other seafoods (Fig. 2). Given the significant bioaccumulation effect of MeHg in the marine food web, there is a correlation between different seafoods[27,28]. Based on this background, this study employed the random forest method to investigate their relationships with the MeHg concentrations in Level 2 and Level 3 marine fishes and utilized these relationships to impute missing values.

Compiling long-term time series national-scale datasets of seafood MeHg concentrations. We used the seafood catch volumes from different marine areas as weights to calculate the national MeHg concentrations of seafoods (Eq. 5). The seafood catch volumes were collected from FAO (Table 1). In this study, we adopted the International Standard Statistical Classification of Aquatic Animals and Plants (ISSCAAP). Specifically, the "Marine Fishes" category was used to represent marine fishes, and the "Crustaceans" and "Molluscs" categories were used to represent other seafoods. Moreover, we classified the marine fish catch volumes based on their trophic levels into the four categories mentioned previously: Level 1 (1.5-2.5), Level 2 (2.5-3.5), Level 3 (3.5-4.5), and Level 4 (4.5-5.5). The seafood catch volumes for certain nations were recorded as zero in certain years. Following Zhang *et al.*[5], we imputed missing seafood MeHg concentrations using the geometric mean of seafood MeHg concentrations from the corresponding continent of each nation. If MeHg concentrations for the whole continent were unavailable, the geometric mean of global MeHg concentration data was used as a replacement.

$$C_{lmt} = \frac{\sum_{j=1}^{n} C_{ljt} * F_{lmjt}}{\sum_{j=1}^{n} F_{lmjt}} \tag{5}$$

The notation C_{lmt} denotes the MeHg concentration of seafood l in nation m at time t; C_{ljt} represents the sample-weighted MeHg concentration of seafood l in marine area j at time t; F_{lmjt} represents the quantity of seafood l caught by nation m in marine area j at time t; and n denotes the number of marine areas.

Data Records

The raw data involved in this study, as well as the compiled MeHg concentration data of seafoods from different marine regions and nations, have been uploaded to Zenodo in XLSX and CSV formats[26]. The raw data are stored in XLSX files named "Raw_species.xlsx", where "species" represents the seafood type, with "FISH" for marine fishes and "OSE" for other seafoods. The processed data used as input for the model are stored in CSV files named "Model_species.csv". Similarly, "species" represents the seafood type, where "FISH" represents marine fishes, and "OSE" represents other seafoods. The capture structures of seafoods are stored in CSV files named "Category_year_capture.csv", where category represents the seafood category, with values including Level 1, Level 2, Level 3, Level 4, and OSE. The MeHg concentrations of seafoods from different marine areas and nations are stored in CSV files named "scale_species_statistics.csv", where "scale" indicates the data scale (i.e., "Marine"

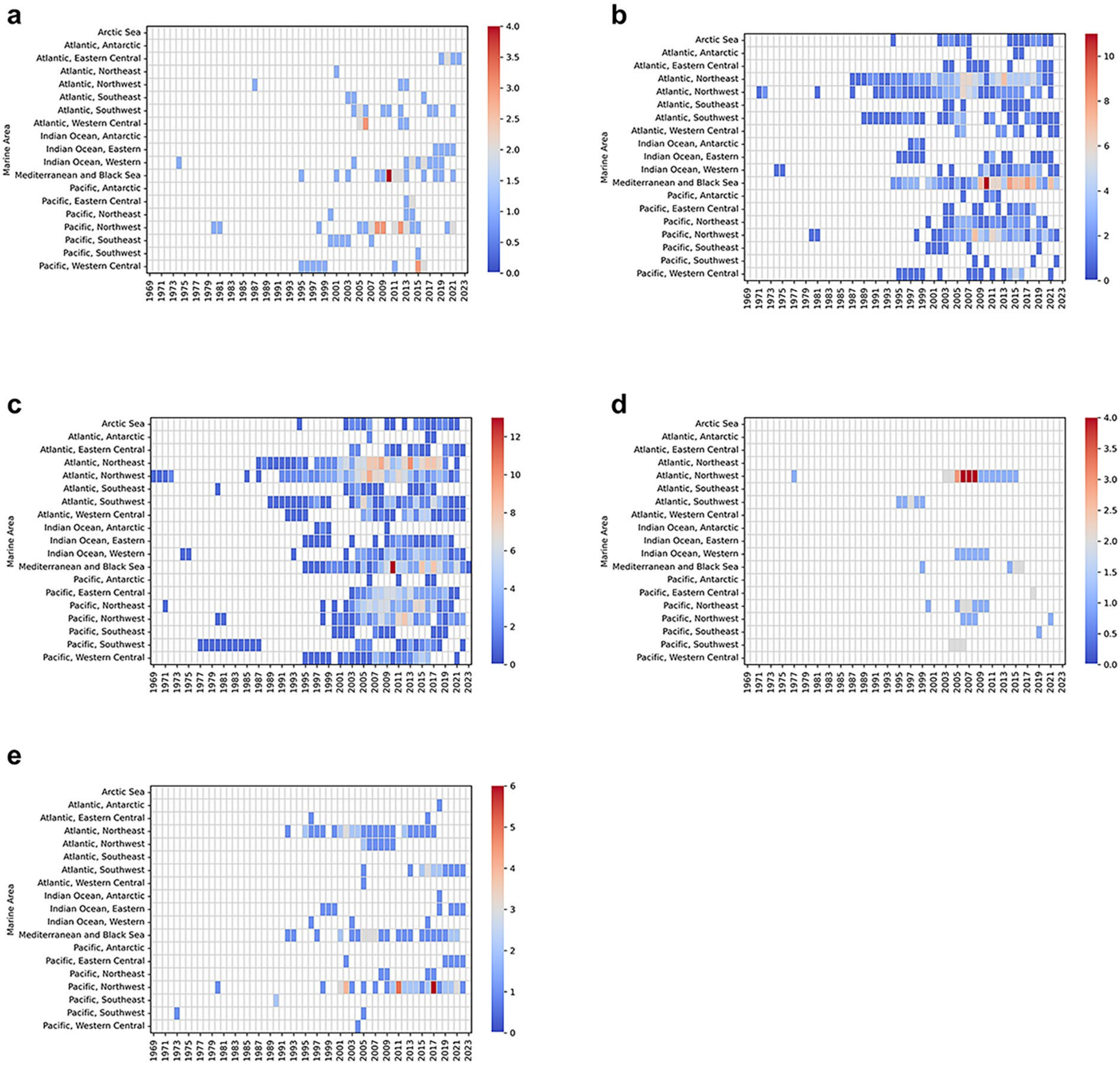

Fig. 2 Temporal distribution of studies on seafood MeHg concentrations in different marine areas. (**a–d**) represent the number of studies on Level 1, Level 2, Level 3 and Level 4 marine fishes respectively, while (**e**) represents the number of studies on seafoods (excluding marine fishes).

for marine regions and "Nation" for countries), "species" represents the seafood category, and "statistics" represents the type of data provided (i.e., "mean" for mean values and "std" for standard deviations of MeHg concentrations). The file "List of literature.xlsx" provides the names of the references from which the concentration data were collected, and "Nation_ID.xlsx" contains the ISO3 codes for each nation. The detailed descriptions of the column parameters in the dataset are provided in Table S1 (see supplementary xlsx file).

Technical Validation

This study conducted a residual analysis to validate the THg to MeHg conversions. The residual analysis for marine fishes demonstrates that the model provides robust predictions of MeHg concentrations. The residual density plot shows a sharp peak centered around zero (Fig. 3a), indicating that most prediction errors are minimal. Moreover, the long but low-density tails on both sides of the distribution reveal that only a small fraction of the predictions exhibit larger errors (Fig. 3a). Overall, the analysis supports the reliability of the model in predicting MeHg concentrations for marine fishes.

For seafoods (excluding marine fishes), the residual analysis also shows that the model predictions are reliable. The residual distribution is centered around zero, indicating that most prediction errors are small (Fig. 3b). Unlike marine fishes, the tails in this distribution are shorter (Fig. 3b), suggesting fewer extreme prediction errors. However, some residuals deviate slightly to the right (Fig. 3b), indicating a potential slight overestimation in certain cases. Overall, the analysis confirms that the model is robust for predicting MeHg concentrations in seafoods (excluding marine fishes).

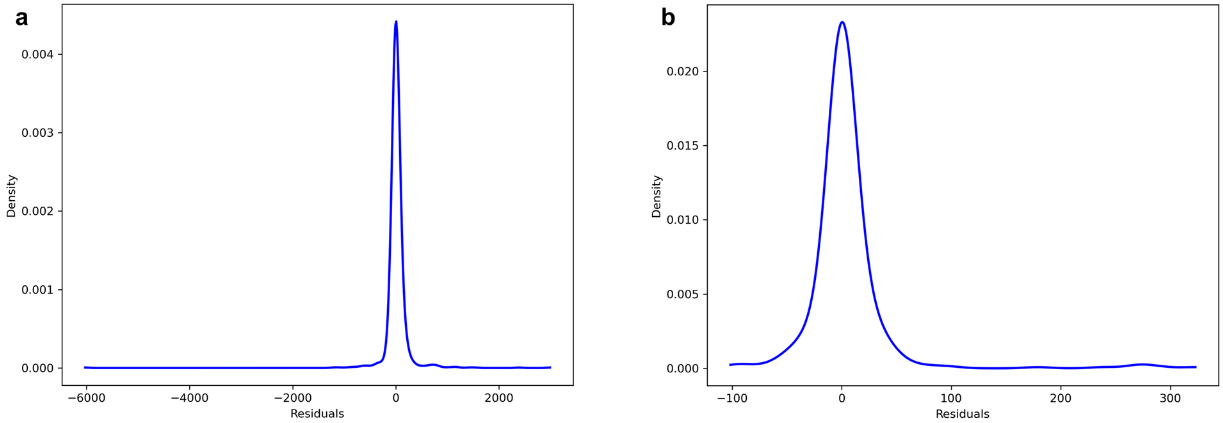

Fig. 3 Kernel density estimate of residuals for model evaluation. (**a**) and (**b**) represent the kernel density estimates of residuals for marine fishes and seafoods (excluding marine fishes), respectively.

This study estimated MeHg concentrations for different trophic levels of marine fishes and seafoods (excluding marine fishes) in various marine areas and nations. However, these estimates cannot be directly compared with species-specific observation data reported in the existing literature. The global seafood MeHg concentration datasets are subject to three uncertainties:

(1) The variation in MeHg concentration measurement methods employed by different studies may introduce methodological biases into data analysis. Specifically, the determination of MeHg requires the combination of Hg determination and Hg speciation analysis. Hg determination methods include, but are not limited to, cold vapour atomic absorption spectrometry (CVAAS), cold vapour atomic fluorescence spectrometry (CVAFS), and inductively coupled plasma-mass spectrometry (ICP-MS), which differ in sensitivity, detection limit, and accuracy[17]. Moreover, Hg speciation analysis techniques, such as gas chromatography (GC), liquid chromatography (LC), and ionic chromatography (IC), also influence the final determination results[19]. These methodological differences may lead to significant discrepancies in reported MeHg concentrations for the same sample across different studies, thereby affecting data integration.
(2) The seafood catch data provided by the FAO does not include non-commercial fisheries, such as subsistence fisheries, recreational fisheries, and illegal fishing, leading to an underestimation of actual catch volumes[29]. As catch is a crucial weighting factor in estimating national MeHg concentrations of seafoods, the lack of data on these fisheries directly affects the accuracy of MeHg concentration estimates.
(3) There were missing values in seafood MeHg concentrations in certain years. This study used random forest algorithms to estimate these missing values. During the model construction process, environmental factors that affect seafood MeHg concentrations (e.g., MeHg concentrations in phytoplankton and seawater, and dissolved organic carbon concentration) could not be incorporated due to the unavailability of long-term time series data for these factors[30,31]. Consequently, the prediction accuracy of the model may be limited due to the lack of multidimensional feature data. Moreover, most existing studies focus on the measurement of THg, while measured data for MeHg remain relatively scarce. This underrepresentation of MeHg measurements introduces further uncertainty into the dataset, as the reliance on THg to MeHg conversions inherently depends on the accuracy of the regression models used.

We used 10,000 Monte Carlo simulations to estimate the uncertainties in seafood MeHg concentrations[32,33]. In the dataset, the uncertainty range of seafood MeHg concentrations was quantified using mean and standard deviation. To reduce the uncertainties presented in the dataset, efforts can be directed toward the following three aspects:

(1) Developing standardized analytical protocols to unify the methods for Hg determination and speciation analysis. This measure will effectively mitigate the determination biases arising from different studies.
(2) Regional collaboration and remote sensing technology should be employed to expand the scope and depth of fisheries data collection. Furthermore, a more comprehensive regional data aggregation mechanism is recommended to encourage nations to provide more complete catch data. These measures will significantly enhance the comprehensiveness and accuracy of fisheries data, thereby improving the precision of seafood MeHg concentration estimates for various nations.
(3) Strengthening the long-term monitoring of relevant environmental factors (e.g., MeHg concentrations in phytoplankton and seawater, and dissolved organic carbon concentrations) and integrating these monitoring data as feature variables into the model can provide more comprehensive input data support for the model. This will contribute to improving the prediction accuracy and reliability of the model. Moreover, efforts should be made to promote the collection and reporting of MeHg concentrations in seafoods. Collaborative initiatives between research institutions and monitoring programs could facilitate MeHg data collection in underrepresented regions, further enhancing the robustness of future datasets.

Usage Notes

The dataset compiled in this study has two potential applications. On one hand, the raw dataset contains THg and MeHg concentrations of global seafoods, supporting a diverse range of research endeavors. For example, this dataset can be used to evaluate the effectiveness of the implementation of the Minamata Convention on Mercury and to identify potential pollution sources. On the other hand, it can provide a data foundation for researchers to assess the potential impacts of MeHg on ecosystems and human health at global and regional scales.

Code availability

The data processing was performed in Microsoft Excel LTSC 16.0. Python 3.11 was used for modeling, and the related codes can be available on Zenodo[26].

Received: 10 May 2024; Accepted: 31 January 2025;
Published online: 11 February 2025

References

1. Grandjean, P., Pichery, C., Bellanger, M. & Budtz-Jørgensen, E. Calculation of Mercury's Effects on Neurodevelopment. *Environ. Health Perspect.* **120**, a452, https://doi.org/10.1289/ehp.1206033 (2012).
2. Roman, H. A. *et al*. Evaluation of the Cardiovascular Effects of Methylmercury Exposures: Current Evidence Supports Development of a Dose–Response Function for Regulatory Benefits Analysis. *Environ. Health Perspect.* **119**, 607–614, https://doi.org/10.1289/ehp.1003012 (2011).
3. Hu, X. F., Lowe, M. & Chan, H. M. Mercury exposure, cardiovascular disease, and mortality: A systematic review and dose-response meta-analysis. *Environ. Res.* **193**, 110538, https://doi.org/10.1016/j.envres.2020.110538 (2021).
4. Zheng, K. *et al*. Epidemiological evidence for the effect of environmental heavy metal exposure on the immune system in children. *Sci. Total Environ.* **868**, 161691, https://doi.org/10.1016/j.scitotenv.2023.161691 (2023).
5. Zhang, Y. *et al*. Global health effects of future atmospheric mercury emissions. *Nat. Commun.* **12**, 3035, https://doi.org/10.1038/s41467-021-23391-7 (2021).
6. United Nations Environment Programme. *Minamata Convention on Mercury*. (UN Environment Programme, Geneva, Switzerland, 2013).
7. Driscoll, C. T., Mason, R. P., Chan, H. M., Jacob, D. J. & Pirrone, N. Mercury as a Global Pollutant: Sources, Pathways, and Effects. *Environ. Sci. Technol.* **47**, 4967–4983, https://doi.org/10.1021/es305071v (2013).
8. Liu, M. *et al*. Impacts of farmed fish consumption and food trade on methylmercury exposure in China. *Environ. Int.* **120**, 333–344, https://doi.org/10.1016/j.envint.2018.08.017 (2018).
9. Karagas Margaret, R. *et al*. Evidence on the Human Health Effects of Low-Level Methylmercury Exposure. *Environ. Health Perspect.* **120**, 799–806, https://doi.org/10.1289/ehp.1104494 (2012).
10. Karimi, R., Fitzgerald, T. P. & Fisher, N. S. A Quantitative Synthesis of Mercury in Commercial Seafood and Implications for Exposure in the United States. *Environ. Health Perspect.* **120**, 1512–1519, https://doi.org/10.1289/ehp.1205122 (2012).
11. Zhang, H. *et al*. Decreasing mercury levels in consumer fish over the three decades of increasing mercury emissions in China. *Eco-Environment & Health* **1**, 46–52, https://doi.org/10.1016/j.eehl.2022.04.002 (2022).
12. Jacobs, S. *et al*. Risk assessment of methylmercury in five European countries considering the national seafood consumption patterns. *Food Chem. Toxicol.* **104**, 26–34, https://doi.org/10.1016/j.fct.2016.10.026 (2017).
13. Chen, L. *et al*. Trans-provincial health impacts of atmospheric mercury emissions in China. *Nat. Commun.* **10**, 1484, https://doi.org/10.1038/s41467-019-09080-6 (2019).
14. Liu, M. *et al*. Rice life cycle-based global mercury biotransport and human methylmercury exposure. *Nat. Commun.* **10**, 5164, https://doi.org/10.1038/s41467-019-13221-2 (2019).
15. Zhou, J. & Obrist, D. Global Mercury Assimilation by Vegetation. *Environ. Sci. Technol.* **55**, 14245–14257, https://doi.org/10.1021/acs.est.1c03530 (2021).
16. Cresson, P. *et al*. Underestimation of chemical contamination in marine fish muscle tissue can be reduced by considering variable wet:dry weight ratios. *Mar. Pollut. Bull.* **123**, 279–285, https://doi.org/10.1016/j.marpolbul.2017.08.046 (2017).
17. Cinnirella, S. *et al*. Mercury concentrations in biota in the Mediterranean Sea, a compilation of 40 years of surveys. *Sci. Data* **6**, 205, https://doi.org/10.1038/s41597-019-0219-y (2019).
18. Sevillano-Morales, J. S., Cejudo-Gómez, M., Ramírez-Ojeda, A. M., Cámara Martos, F. & Moreno-Rojas, R. Risk profile of methylmercury in seafood. *Curr. Opin. Food Sci.* **6**, 53–60, https://doi.org/10.1016/j.cofs.2016.01.003 (2015).
19. Zmozinski, A. V. *et al*. Method development for the simultaneous determination of methylmercury and inorganic mercury in seafood. *Food Control* **46**, 351–359, https://doi.org/10.1016/j.foodcont.2014.05.054 (2014).
20. Jinadasa, B. K. K. K., Jayasinghe, G. D. T. M., Pohl, P. & Fowler, S. W. Mitigating the impact of mercury contaminants in fish and other seafood—A review. *Mar. Pollut. Bull.* **171**, 112710, https://doi.org/10.1016/j.marpolbul.2021.112710 (2021).
21. Sadhu, A. K., Kim, J. P., Furrell, H. & Bostock, B. Methyl mercury concentrations in edible fish and shellfish from Dunedin, and other regions around the South Island, New Zealand. *Mar. Pollut. Bull.* **101**, 386–390, https://doi.org/10.1016/j.marpolbul.2015.10.013 (2015).
22. Zhang, Y. *et al*. Mapping Blood Lead Levels in China during 1980–2040 with Machine Learning. *Environ. Sci. Technol.* **58**, 7270–7278, https://doi.org/10.1021/acs.est.3c09788 (2024).
23. Wang, Y. *et al*. A Random Forest Model for Daily $PM_{2.5}$ Personal Exposure Assessment for a Chinese Cohort. *Environ. Sci. Technol. Lett.* **9**, 466–472, https://doi.org/10.1021/acs.estlett.1c00970 (2022).
24. Brokamp, C., Jandarov, R., Hossain, M. & Ryan, P. Predicting Daily Urban Fine Particulate Matter Concentrations Using a Random Forest Model. *Environ. Sci. Technol.* **52**, 4173–4179, https://doi.org/10.1021/acs.est.7b05381 (2018).
25. Liu, Y.-R. *et al*. Multidimensional Drivers of Mercury Distribution in Global Surface Soils: Insights from a Global Standardized Field Survey. *Environ. Sci. Technol.* **57**, 12442–12452, https://doi.org/10.1021/acs.est.3c04313 (2023).
26. Zhou, H., Li, Y., Zhong, Q., Wu, X. & Liang, S. A global seafood methylmercury concentration dataset during 1995-2022. *Zenodo* https://doi.org/10.5281/zenodo.14607175 (2025).
27. Vainio, R. K. *et al*. Trophic Dynamics of Mercury in the Baltic Archipelago Sea Food Web: The Impact of Ecological and Ecophysiological Traits. *Environ. Sci. Technol.* **56**, 11440–11448, https://doi.org/10.1021/acs.est.2c03846 (2022).
28. Wu, P. & Zhang, Y. Toward a Global Model of Methylmercury Biomagnification in Marine Food Webs: Trophic Dynamics and Implications for Human Exposure. *Environ. Sci. Technol.* **57**, 6563–6572, https://doi.org/10.1021/acs.est.3c01299 (2023).
29. Belhabib, D., Koutob, V., Sall, A., Lam, V. W. Y. & Pauly, D. Fisheries catch misreporting and its implications: The case of Senegal. *Fish. Res.* **151**, 1–11, https://doi.org/10.1016/j.fishres.2013.12.006 (2014).
30. Wu, P., Dutkiewicz, S., Monier, E. & Zhang, Y. Bottom-Heavy Trophic Pyramids Impair Methylmercury Biomagnification in the Marine Plankton Ecosystems. *Environ. Sci. Technol.* **55**, 15476–15483, https://doi.org/10.1021/acs.est.1c04083 (2021).

31. Lavoie, R. A., Jardine, T. D., Chumchal, M. M., Kidd, K. A. & Campbell, L. M. Biomagnification of Mercury in Aquatic Food Webs: A Worldwide Meta-Analysis. *Environ. Sci. Technol.* **47**, 13385–13394, https://doi.org/10.1021/es403103t (2013).
32. Zhang, X., Zhong, Q., Chang, W., Li, H. & Liang, S. A high spatial resolution dataset for methylmercury exposure in Guangdong-Hong Kong-Macao Greater Bay Area. *Sci. Data* **10**, 706, https://doi.org/10.1021/es403103t (2023).
33. Lu, H. *et al*. Greenhouse gas emissions from U.S. crude oil pipeline accidents: 1968 to 2020. *Sci. Data* **10**, 563, https://doi.org/10.1038/s41597-023-02478-4 (2023).

Acknowledgements
This study was financially supported by the National Natural Science Foundation of China (Grant No. 52325005, 72293602, and 52388101).

Author contributions
H.Z. calculated and verified the dataset. S.L. led the project and provided the methods. H.Z., Y.L. and Q.Z. collected the raw data and programmed the calculation steps. X.W. checked the articles and data. All authors contributed to the writing and revisions of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-04570-3.

Correspondence and requests for materials should be addressed to S.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025