Mixed effects model approaches for estimating size-adjusted contaminant concentrations in fish populations.

\* Emily Smenderovac (corresponding author), Great Lakes Forestry Centre, Natural Resources Canada, [emily.smenderovac@nrcan-rncan.gc.ca](mailto:emily.smenderovac@nrcan-rncan.gc.ca)

Brian W. Kielstra, Great Lakes Forestry Centre, Natural Resources Canada, [bkielstra@ecometrix.ca](mailto:bkielstra@ecometrix.ca)

Tom Johnston, Living with Lakes Centre, Laurentian University, Ontario Ministry of Natural Resources and Forestry, [tjohnston@laurentian.ca](mailto:tjohnston@laurentian.ca)

Satyendra Bhavsar, Ontario Ministry of the Environment and Climate Change, [satyendra.bhavsar@ontario.ca](mailto:satyendra.bhavsar@ontario.ca)

Rob Mackereth, Centre of Excellence for Sustainable Mining and Exploration, Ontario Ministry of Natural Resources and Forestry, [rob.mackereth@ontario.ca](mailto:rob.mackereth@ontario.ca)

Stephanie Melles, Department of Chemistry and Biology, Toronto Metropolitan University, [stephanie.melles@torontomu.ca](mailto:stephanie.melles@torontomu.ca)

Gretchen L. Lescord, Forests Fisheries and Geomatic Sciences, University of Florida, [lescord.g@ufl.edu](mailto:lescord.g@ufl.edu)

Erik J.S. Emilson, Great Lakes Forestry Centre, Natural Resources Canada, [erik.emilson@nrcan-rncan.gc.ca](mailto:erik.emilson@nrcan-rncan.gc.ca)

2023-10-26

# 1 Abstract

Larger fish are known to have higher concentrations of bioaccumulative contaminants like mercury (Hg). Sampling event regression (SERs) are frequently used to estimate size-adjusted concentrations of these contaminants in fish populations prior to making comparisons or further modeling. However, this approach can be limited by sample size constraints within and across waterbodies. Herein, we compare mixed effect linear models (LMER), Bayesian inference (RSTAN) and approximate Bayesian inference (INLA) approaches that borrow strength from all available observations to estimate size-adjusted contaminant concentrations in individual fish populations. We found that RSTAN and INLA produced comparable predictions to SER approaches, with increased accuracy and positive bias for estimation errors. LMER models conversely, had less accuracy and error trended towards underestimating values. The INLA and RSTAN approach allowed for the estimation of values for lake populations with less than 5 fish, and which lacked the range of fish sizes required for SER estimates. INLA is the more suitable approach, with fast speeds and better overall accuracy than RSTAN models. We recommend INLA as a reliable means of estimating size-adjusted contaminant concentration in fish populations, particularly when datasets include some lake populations where sample sizes are limited.

# 2 Key Words

Mercury, Arsenic, Fish, Mixed effects models, contaminant modelling

# 3 Synopsis

!!! TOC art will go here !!!

# 4 Introduction

Consumption of wild-caught fish, while abundant with healthy fats and protein, is considered a potentially detrimental source of contaminants to the human body1. Bioaccumulative contaminants - which are incorporated into tissues faster than they can be excreted - are of particular concern because they can reach elevated concentration in fish, even in remote waterbodies distant from primary sources of pollution. These contaminants tend to biaccumulate as the fish ages, so concentrations are often correlated with metrics of fish body size (i.e., weight and length). As a result, many monitoring programs and studies account for fish weight or length in their consumption recommendations (e.g.,2,3). This accumulation-size relationship is also accounted for in statistical models used in studies of contaminants to avoid confounding effects of fish size4,5. A commonly used approach for estimating size-adjusted means of contaminant concentrations at the population-level is regressing contaminant concentrations against metrics of body size within a fish population (usually a unique species-lake combination for a given year) and use the resulting linear equation, or an ANCOVA model to estimate the concentration at a chosen size (e.g., 500 g or 1kg)4–6; herein we refer to this approach as Sampling Event Regressions (SERs).

The strength of fish contaminant-size relationships vary among different types of contaminants and across fish species or waterbodies, due to complex environmental and metabolic factors. For example, mercury (Hg) is a highly bioaccumulative contaminant that, due to its ubiquitous dispersion, is a global concern that is routinely monitored in fish. Mercury is well known to accumulate in the muscle tissue of fish, due its chemical speciation and inter-organ transport7. Many studies have found strong positive correlations between Hg concentrations ([Hg]) and body size in both freshwater8 and marine9 environments and, thus, size-adjustments of Hg concentrations prior to making comparisons across environmental gradients are common5,6,10–12. In contrast, arsenic is less bioaccumulative in fish, showing mixed and weaker relationships with metrics of body size5,13. Depending on its chemical speciation, arsenic can be a harmful carcinogenic contaminant, but it is less widely distributed and often originates from a localized point source of contamination or natural abundance. It is also less routinely monitored and researched when compared to [Hg] in fish, which results in less data to be available for size-adjustment models. However, not accounting for size in population-level comparisons of arsenic concentrations ([As]) could mean that some variation is unaccounted for, possibly weakening predictions and limiting our understanding of environmental cycling and bioaccumulation patterns.

In general, research and monitoring programs sample waterbodies for fish seek to attain minimum sample size (e.g., 10-20 individuals/species) across as broad of a size range as possible to ensure a reasonable estimate of the population8,14,15. However, opportunistic sampling can limit the number of fish caught, the size range represented, or the distribution of fish sizes for a given population – all of which may lead to over/under estimating standardized concentrations when using an SER approach. Small sample sizes can be particularly limiting in assessments of temporal trends, and may lead researchers to pool samples over time periods like in Bhavsar *et al.,*16. Alternatively, deficient populations may be excluded altogether, sacrificing valuable data often from underrepresented waterbodies or fish species. Mixed effects regression approaches have been used to overcome this limitation such as maximum likelihood mixed effects methods17. Bayesian linear mixed effects models18 are also potential solution as they utilize all data across sampling structures (e.g., waterbodies) to inform predictions and allow for explicitly modeling variation of random effects variables. Different Bayesian linear mixed effects models have been implemented in R, including Bayesian inference in the (rstanarm) package19, and approximate Bayesian inference in the (INLA) package20,21.

Herein, we present a novel statistical approach to generating size-adjusted contaminant estimates at the population level using approximate Bayesian inference (INLA) and compare the results to SER, mixed effects regression (using the lmer package) and Bayesian inference (RSTAN). We evaluate these approaches using two contaminant measures, total [Hg] and [As] in fish muscle tissue, to assess differences in the modeling approaches based on bioaccumulative potential and data availability. More specifically, we compared the results of mixed model approaches that borrow strength from fish across all populations (i.e., pooled across waterbodies and through time) to generate predictions for individual sampling events on lakes to the SER approach. We hypothesize that mixed modelling approaches will provide comparable or more accurate predictions of contaminant concentrations to SER. We also expect that Bayesian approaches will have improved accuracy compared to more simplistic implementations of linear mixed models, especially for predictions in lakes with low sample numbers. We expect that INLA will outperform STAN models in terms of computation time, but have similar predictive accuracy.

# 5 Materials and Methods

## 5.1 Datasets

We obtained fish-level Hg and As data for inland waterbodies across Ontario from the Ontario Ministry of Environment, Conservation, and Parks (OMECP). Although individual fish sampling protocols varied, muscle tissue sent to OMECP is analyzed using standardized methods as part of the Fish Contaminants Monitoring Program. More specifically, total [Hg] were measured using cold vapor-flameless atomic absorption spectroscopy (CV-FAAS) following protocol HGBIO-WS057 and total [As] was measured using Inductively Coupled Plasma Mass Spectrometry (ICP-MS) following method BIOTA-E3461. We limited our analyses to inland lakes and three species of interest: Salvelinus namaycush namaycush (common Lake Trout, hereafter LT), Esox lucius (Northern Pike, hereafter NP), and Sander vitreus (Walleye, hereafter WE). These three species represent important food fish in Ontario, sought by subsistence and sport fishers across the province. While they are all predatory fish, LT are generally restricted to profundal zones, while NP and WAL have broader movement patterns.

## 5.2 Data Characterization

In total, the final dataset we used included [Hg] in 37923 fish and [As] in 1001 fish. We decided on two weight-standardization targets of 500g and 1000g, as these were well represented in the dataset11.2, and have been used in other studies5,6,22. Fish from the targeted weight classes were randomly sampled to create a test dataset with ~50% of the fish within 400 - 600g and 900 - 1100g and a train dataset with the remaining data from other fish size classes, as well as the remaining ~50% of fish in the targeted size classes (Tables 5.1, 5.2)

Table 5.1: Summary of events (unique year and lake combinations) for each contaminant and species in the train dataset used in this analysis.

| CONTAMINANT | SPECIES\_NAME | Total number of events | Events with more than one fish | Events with at least 5 fish with proper weight range | Number of events excluded from SER | Number of lakes excluded from SER | Total number of unique lakes |
| --- | --- | --- | --- | --- | --- | --- | --- |
| As | Lake Trout | 27 | 24 | 10 | 9 | 14 | 22 |
| As | Northern Pike | 102 | 95 | 19 | 29 | 72 | 89 |
| As | Walleye | 97 | 87 | 21 | 26 | 63 | 82 |
| Hg | Lake Trout | 637 | 602 | 396 | 92 | 129 | 411 |
| Hg | Northern Pike | 1469 | 1428 | 1017 | 160 | 251 | 999 |
| Hg | Walleye | 1441 | 1408 | 1172 | 99 | 134 | 951 |

Table 5.2: Summary of events (unique year and lake combinations) for each contaminant and species in the test dataset used in this analysis.

| CONTAMINANT | SPECIES\_NAME | Total number of events in prediction sample | Events modelled with more than one fish | Events modelled with at least 5 fish | Number of event predictions excluded from SER | Number of lakes predictions excluded from SER | Total number of unique lakes included in predictions |
| --- | --- | --- | --- | --- | --- | --- | --- |
| As | Lake Trout | 7 | 6 | 2 | 2 | 2 | 5 |
| As | Northern Pike | 35 | 31 | 8 | 25 | 24 | 33 |
| As | Walleye | 30 | 27 | 8 | 20 | 19 | 28 |
| Hg | Lake Trout | 313 | 309 | 273 | 40 | 38 | 243 |
| Hg | Northern Pike | 859 | 840 | 737 | 105 | 96 | 664 |
| Hg | Walleye | 1047 | 1036 | 937 | 66 | 65 | 765 |

## 5.3 Modelling

All modelling and analyses were performed in R Version 4.2.323 on a Dell Latitude 5510 PC running Windows 10 Enterprise with a 1.70GHz, 2208 MHz 4 Core(s), 8 logical Processor, 16 GB of physical memory and 34.6 GB of virtual memory. All data transformations, summarizations and graphing were performed with the tidyverse, ggplot2 and ggpubr packages24–26.

### 5.3.1 Sampling event regressions

For each contaminant-species-waterbody-year combination (i.e., a sampling event), we developed log-contaminant (ug/g) by log-weight (g) regression models for those combinations with at least 5 sampled individuals, and only for models where there were values above and below at least one of our targeted weights for standardization. Each model used a linear equation (Equation (5.1).

where *i* is an individual fish for a species-waterbody-year combination; these models were run on training data until all combinations were exhausted. The time required to run each model was recorded, and we used these models to generate contaminant predictions and their 95% confidence intervals for 500g and 1000 g fish per sampling event, as well as predict contaminant concentrations of testing data. Predicted values of test data were collected and used to generate root mean squared error for each sampling event model (RMSEevent), effectively representing the predictive accuracy of each model.

### 5.3.2 Mixed effects regression models

For each contaminant-species combination, we developed log-contaminant (μg/g) by log-weight (g) mixed effects regression models. We allowed for random variation in the slope and intercept per waterbody, and random variation in the intercept by sampling event (i.e., each unique waterbody-year combination). This identical model structure was fit in three modelling approaches: maximum likelihood inference in lme4 (LMER and boot\_LMER models;27), Bayesian inference with Markov Chain Monte Carlo in Stan through rstanarm (STAN models;19), and approximate Bayesian inference using integrated nested Laplace approximation in INLA through R-INLA (INLA models;20). These models used a random mixed effects formula (Equation (5.2)).

Where *U* is the random effect on the slope for each waterbody, and *W* is the random effect of each waterbody-year sampling event.

For each modelling approach, the time required to run each model was recorded, and the residuals of the testing dataset were collected to calculate RMSEevent for the distinct sampling events in the dataset. We used these models to generate contaminant predictions and their 95% confidence intervals for 500g and 1000 g fish per sampling event.

#### 5.3.2.1 Maximum likelihood inference

ML models were fit using the (lme4) package, using default settings27. We generated results for the basic model, with the understanding that this model would not incorporate the uncertainty of the random effects. To account for this limitation, we also performed a bootstrapping analysis using (lme4)’s (bootMer) parametric bootstrapping with effects using 2000 simulations and with the use.u setting set to TRUE using 4 cpus. We used the bootstrapping results to generate 95% confidence intervals of the ML coefficients and predictions that would include the uncertainty in the random effects.

#### 5.3.2.2 Bayesian inference with Markov Chain Monte Carlo

For STAN models, all models were run with chains and cores set to 4. We ran the models for [Hg] datasets using the RSTAN default settings19. [As] were lower sample size, and so required higher iterations to achieve reliable model results, these were run with 9000 iterations and an adapt\_delta setting of 0.99, this essentially makes acceptance criteria for posterior distributions during the adaptation period more strict, and lowers the step-size of the model, while increasing the computation time. The 50th percentile of predicted values were used to calculate residuals, RMSEevent and RMSEglobal for these models.

#### 5.3.2.3 Approximate Bayesian inference using integrated nested Laplace approximation

INLA models were run in the R-INLA package, with precision of priors set to 0.001, and the formula structure was entered for the “iid2d” model to account for the covariance of the random effects due to waterbody20. The 50th percentile of predicted values were used to calculate residuals, RMSEevent and RMSEglobal for these models.

## 5.4 Comparing accuracy of predictions across model types

We compared model residuals of fish in 500 g and 1000 g size ranges (as calculated by the actual test values subtracted from the predicted values) to visualize the range in predictive error for each approach. The RMSEevent for each of the modelling approaches, as well as the RMSEglobal of each model against the runtime required for each approach was based off the fit of each model to the training data. We also compared predicted values for 500g and 1000g fish for each sampling event in the dataset. We used the wilcox.test function R to perform wilcoxon two-sided tests to evaluate the deviation of residuals from zero, and we compared predicted values to true values, and SER predicted values using Pearson correlation with the cor.test function in R, and visually assessed the overlap of the linear relationships with a 1:1 relationship with 25% error.

# 6 Results

## 6.1 Model implementation

Application of each approach varied in difficulty, going from LMER<SER<STAN<INLA. The LMER approach was simple enough that any beginner R user could implement it quickly, without requiring knowledge of loops or other programming. The SER approach required more programming knowledge in order to implement it efficiently, and a significant amount of pre-screening of data to ensure that data used for linear models covered a large enough span to develop a curve. There were more documentation and reading requirements to implement the RSTAN and INLA models. There is a higher risk of incorrect implementation of INLA models with the current INLA package implementation, since the model formula input deviates from the typical R notation. There is thorough documentation of the model notation in the INLA package, but it is nonetheless a consideration in the implementation of this approach, as it added complexity when assessing the nested random effect of waterbody and sampling event. The [As] models were generally more difficult to fit due to lower sample sizes, requiring different settings, while [Hg] models ran well on the default settings of the lmer, rstan and inla functions.

## 6.2 Data biases

The dataset was heavily biased to smaller fish, but had a normal distribution of values around the targeted weights, which were highly represented in the data (Supplemental Fig 11.2). Generally, the dataset was dominated by low levels of contaminant, less than 1.3 ug/gram~wet weight~ [Hg] and less than 0.4 ug/g~wet weight~ [As] (Supplemental Fig 11.1).

## 6.3 Prediction results

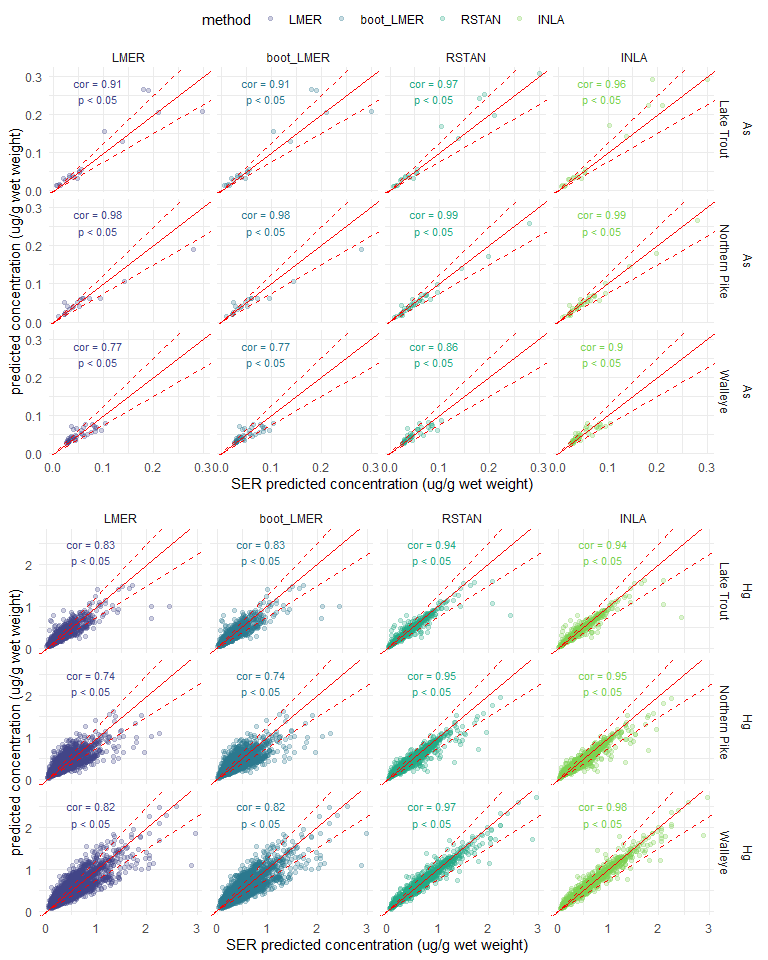


Figure 6.1: Predicted concentration of [As] and [Hg] in fish tissue of 400-600g and 900-1100g fish compared to known values. Points display each individual fish prediction and a solid red line shows a 1-1 relationship. Dashed red lines provide an interval for predicted values that are within 25% of the predicted SER value. Correlation statistics are included in the top left of each panel.

Predictions from LMER and LMER-bootstrapped approaches were consistent with each other, and had a general association with predicted values from SER, but much more noise in the predictions than the RSTAN or INLA approaches. The RSTAN and INLA approaches both had predicted values that performed very comparably with SER predictions, with most values greater than 1 ug/g~wet weight~ falling within 25% of the predicted SER values. The values lower than 1 ug/g~wet weight~ had more variance than this, but stayed within approximately 0.5 ug/g~wet weight~(Fig 6.1). Additional graphing of these predictions revealed that much of the deviations in these predictions were from predictions of the 1000g weights, and the 500g weight predictions had stronger associations (Supplemental Figs 11.3, 11.4).

## 6.4 Model performance by sampling event

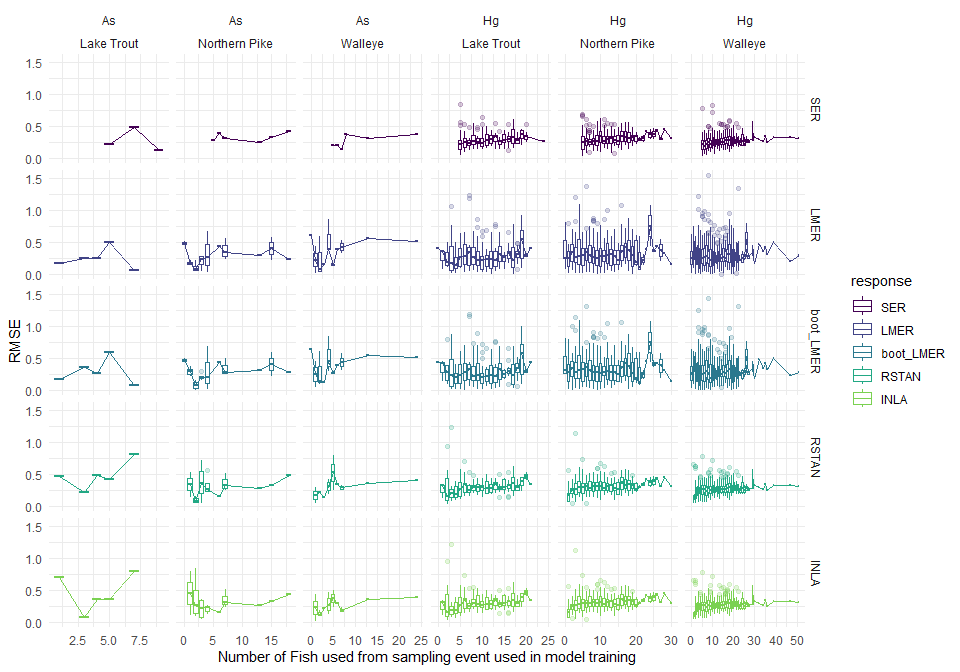


Figure 6.2: Residual mean squared error of models from sampling events trained with different numbers of fish. RMSE of training set sampling events are compared to the number of fish used from the sampling event that were included in the training dataset. For each N, a boxplot displays the 25th and 75th percentiles and the median RMSE, with the whiskers extending out to 1.5 x the interquartile range. A line is plotted across the median values to show the overall trend in the values.

All approaches had RMSE for events that generally increased and stabilized at about 6 fish sampling events. LMER and LMER bootstrapped approaches had the highest RMSE ranges in [Hg] datasets (0 - 1.55) while SER, RSTAN and INLA approaches had comparable fits, with a lower range of RMSEevent that were more consistently between 0.03 - 1.24 RMSE. There was some additional noise, and higher RMSEevent for [Hg] and [As] RSTAN and INLA models compared to the SER approach, but much of that was in the lower N lakes. RMSE for As models were more consistent between approaches with a range of 0 - 0.86 RMSEevent (Fig 6.2).

## 6.5 Error distribution amongst models

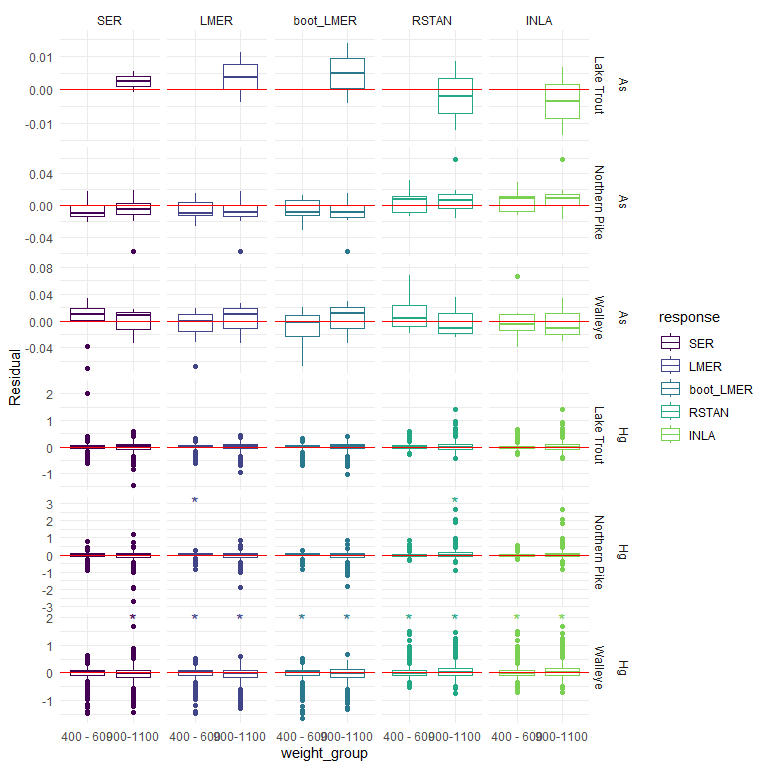


Figure 6.3: Residuals for fish from events modelled with at least 5 fish, within 100 g of the targeted prediction weights (500g and 1000g) from each modelling approach in the study. Values are displayed in boxplots where the edges of the box represent the 25th and 75th percentiles and the central line the median, with the whiskers extending out to 1.5 x the interquartile range. A red horizontal line on each plot indicates where there is a residual of zero. ’\*’ are used to represent medians that significnatly deviate from zero based on two sided wilcoxon tests.

Residuals of testing data for high-N (at least 5 fish) sampling events showed that for all approaches, the median value of [Hg] residuals for fish in the range of the standardized weights of interest (500 g and 1000 g) generally trended towards zero, though had some small but significant (p < 0.05) differences from medians For Walleye models, for most approaches, for Northern Pike 400-600 g LMER, bootstrapped LMER and INLA approaches. While medians of [As] values for all approaches visually deviated from zero none of these deviations were significant. The outliers in these data indicated a potential bias for overestimation of some values in the RSTAN and INLA approaches, while more of the outliers for the SER, LMER and bootstrapped LMER approaches were underestimated (Fig 6.3). Bootstrapped LMER showed little difference from LMER values.

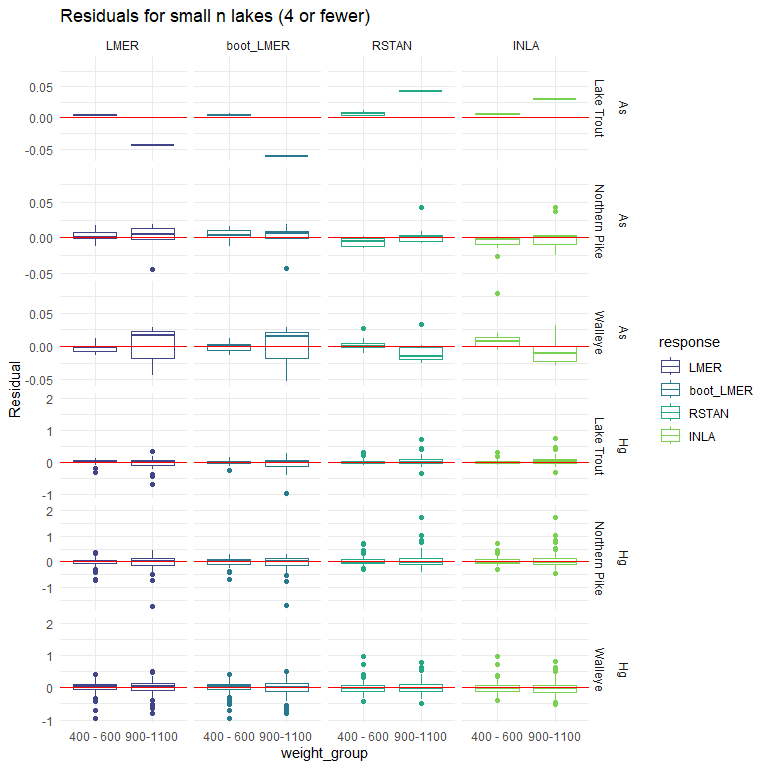


Figure 6.4: Residuals for fish from events modelled with less than five fish, within 100 g of the targeted prediction weights (500g and 1000g) from each modelling approach in the study. Values are displayed in boxplots where the edges of the box represent the 25th and 75th percentiles and the central line the median, with the whiskers extending out to 1.5 x the interquartile range. A red horizontal line on each plot indicates where there is a residual of zero. ’\*’ are used to represent medians that significnatly deviate from zero based on two sided wilcoxon tests.

When only the small-N sampling events (4 or less fish) were investigated, no significant (p < 0.05) differences from zero were present, and there was visual overlap of medians and the zero line, indicating that the LMER, bootstrapped LMER, RSTAN and INLA approaches were estimating contaminant values for those lakes as effectively as the lakes that could be modelled using SER (Fig 6.4).

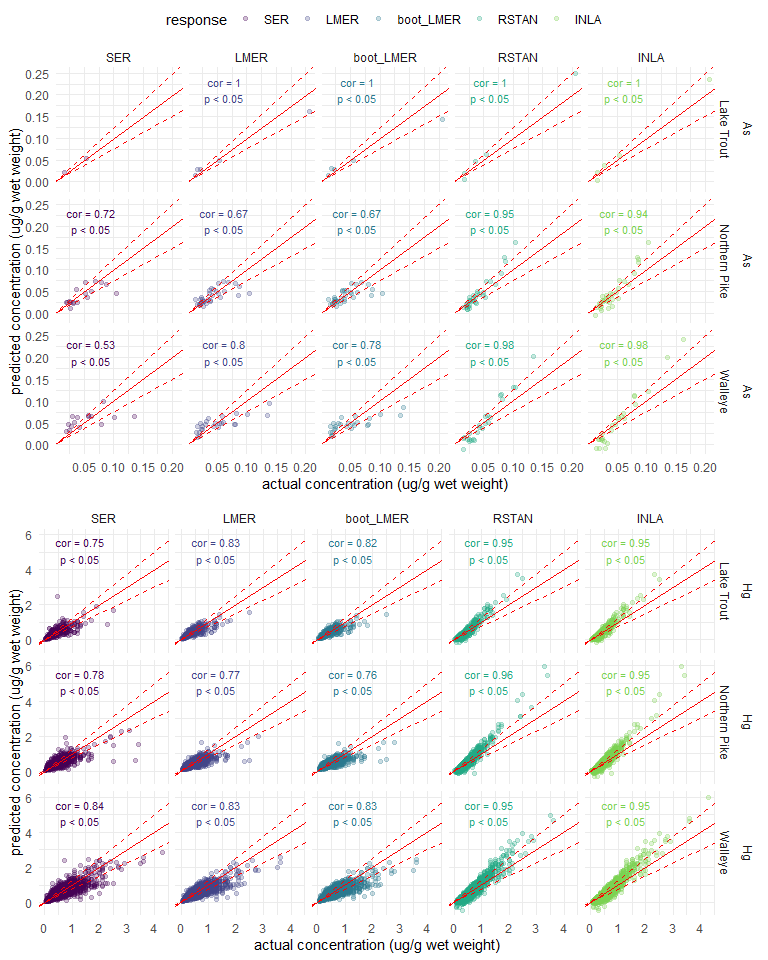


Figure 6.5: Predicted concentration of [As] and [Hg] in fish tissue of 400-600g and 900-1100g fish compared to known values. Points display each individual fish prediction and a solid red line shows a 1-1 relationship. Dashed red lines provide an interval for predicted values that are within 25% of the measured value. Correlation statistics are included where there were enough values to compute them.

Each modelling approach resulted in different error distributions, with all approaches estimating test data values with averaged accuracy for [As] fallig close to to 25% accuracy at most concentrations, and good accuracy (average estimated prediction within 25% accuracy) for [Hg] models up to a concentration of approximately 1.25 ug/gwet weight. There were biases in observed error that appeared to be driven by test values higher than 1.25 ug/gwet weight, with SER, LMER and bootstrapped LMER values resulting in underestimated predictions, and RSTAN and INLA resulting in overestimated values (Fig 6.5).

## 6.6 Relative computational requirements

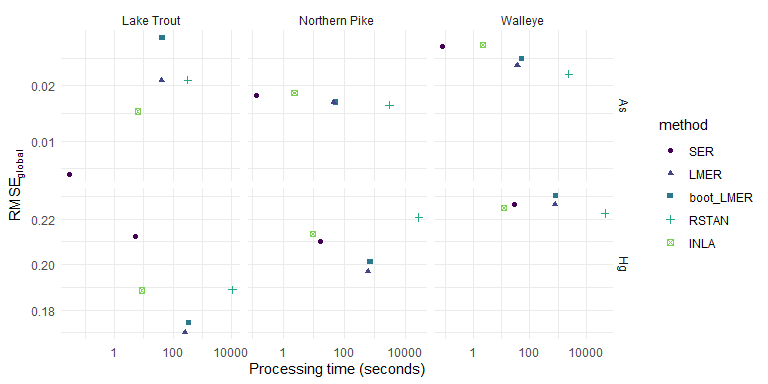


Figure 6.6: Global RMSE for each modelling approach compared to the computation time required to perform it. The time in seconds, is presented on a log-scaled axis.

Generally, the LMER, RSTAN and INLA models had lower global RMSE and better model fit than the SER approach, with the exception of the [As] Lake trout model, which had the lowest sample size of all models in this study. While SER was generally the quickest of all modelling approaches for [As] datasets, INLA had comparable speeds in the larger [Hg] datasets. LMER and LMER bootstrapping had variable global RMSE, but they were generally not the lowest RMSE for any modelling set, aside from the [Hg] Lake Trout model, where they performed comparatively with the INLA model. LMER approaches were more time intensive than either INLA or SER, and bootstrapping did not reliably reduce RMSE. RSTAN also had variable performance in terms of global RMSE and took the most computational time of any approach, though it generally had lower error compared to SER, it resulted in higher error compared to the LMER models in the [Hg] Lake Trout dataset and [As] Northern Pike and Walleye datasets, and the SER approach in the [As] Lake Trout dataset. The INLA models had the lowest computational times and RMSE relative to all other models, with the exception of the [As] Lake Trout dataset (Fig 6.6).

# 7 Discussion

The Bayesian approaches resulted in the most comparable predictions to the SER approach, with the added benefit of adding predictions for low sample number lake sampling events (less than 5 fish) that were comparable to the accuracy of predictions from higher sample number lake sampling events (5 fish or more). The LMER approach had more noise in the relationship with SER predictions, demonstrated by lower R2 fits of the relationship between values predicted by the two methods. LMER, bootstrapped LMER, RSTAN and INLA approaches had an approximately linear relationship with SER predictions. It is likely that there was more error for higher values of contaminant in the dataset, as these values were underrepresented in the data used for training these models compared to lower levels of contamination. This seems to be a naturally occurring characteristic of the dataset (there are simply not as many lakes and fish with high levels of contaminant) it will continue to be a challenge for accurate standardization of values for higher contaminant levels, and introduces some bias into the models. In freshwater systems, [Hg] typically ranges from …. for Lake Trout, … for Northern Pike and … for Walleye, which align quite well with the range we observed in the data …. for Lake Trout, … for Northern Pike and … for Walleye. The higher variance in response due to LMER approaches is likely due to the limited capability for this approach to account for uncertainty in random effects, and it appears that our bootstrapping implementation did little to correct for this issue.

While the accuracy was not consistent across the contaminant gradient, Bayesian approaches were able to predict [As] and [Hg] concentrations in fish tissue to comaparable levels of the SER and lmer approaches, and achieve this level of accuracy in lakes sampling events with less than 5 fish across all three fish types. We conclude INLA and RSTAN can standardize contaminant concentrations from events with less than 5 fish to our 500g and 1000g targets as effectively as higher-saple number events in the SER approach. As expected, all models performed better for sampling events with more fish, and the biases in estimation that appeared in the different datasets correlated with differences in representation of values in the underlying dataset. Future work will aim to determine whether Bayesian models with multiple fish species in the same model can overcome some of the limitations related to bias in the distribution on contaminant concentration, or whether differences in lifestyle and ecology introduce more noise and reduce predictive accuracy.

While both Bayesian approaches performed comparably for most metrics, INLA models were much less computationally intensive and had improved RMSEglobal compared to the RSTAN implementation. As these approaches were comparable in terms of difficulty of implementation, and the accuracy of the INLA approach is improved, we suggest that INLA is a viable alternative to standardizing values via the SER approach, that may allow for incorporation of more data into analyses.

Each had different benefits and drawbacks in terms of implementation. Conceptually, the SER approach is the most simple to understand, given that is essentially a group of independent linear models. This can also mean there is more manual intervention required, as error is not ‘averaged out’ over the dataset. The LMER, STAN and INLA are all more conceptually challenging, however, there is quite a jump in the complexity of a Bayesian model from the LMER approach. Comparatively, understanding of the Bayesian (RSTAN), and the approximated Bayesian (INLA) approach are again more complex, which should be considered when model tuning is required, which based on this study, would likely be required for smaller datasets (<300).

The different approaches resulted in different biases in predictions. While SER and LMER approaches were more likely to underestimate high values, based on results from testing data, the Bayesian approaches appeared to be more likely to overestimate high values. This conflicted with the comparison of predicted results, but this was likely a result of higher error associated with SER and LMER estimates, which was demonstrated with the lower R2 of these models than the RSTAN and INLA approaches. All approaches had similar losses in accuracy as the amount of contaminant increased, due to underepresentation of high contaminant values in the underlying dataset. The tendency of the INLA approach to overestimate values may be a more desirable manifestation of estimation error for developing fish consumption guidelines and protection of fish stocks, but may be result in overly cautious prescriptions.

Results of the modelling approaches does not suggest that the different lifestyles of the assessed fish species influence the performance of the model, so this approach is likely to be suitable for many types of fishes.

# 8 Conclusion

# 9 Acknowledgement

# 10 Data Accessibility

The processed data and scripts used to create this manuscript are available for download at the public repository managed by the WETlab at the Great Lakes Forestry Centre: <https://github.com/GLFC-WET/HGAS_master>

# 11 Supplemental Figures

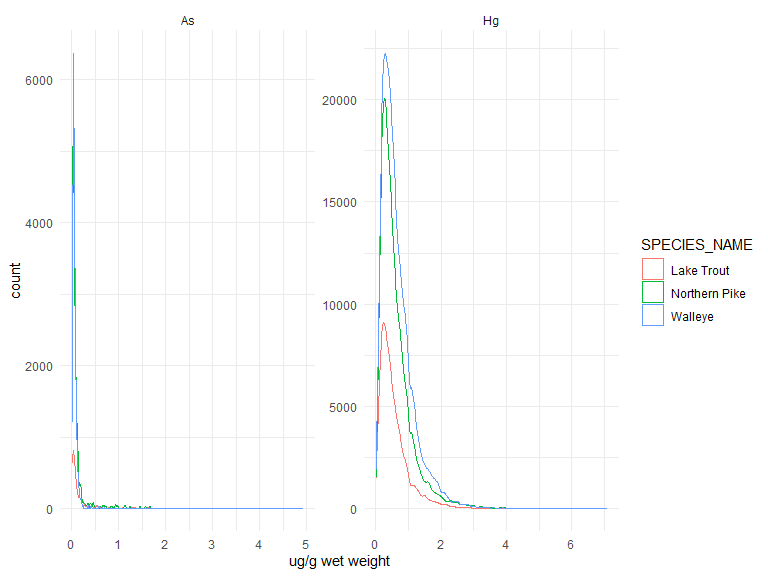


Figure 11.1: Number of fish with different contaminant concentrations

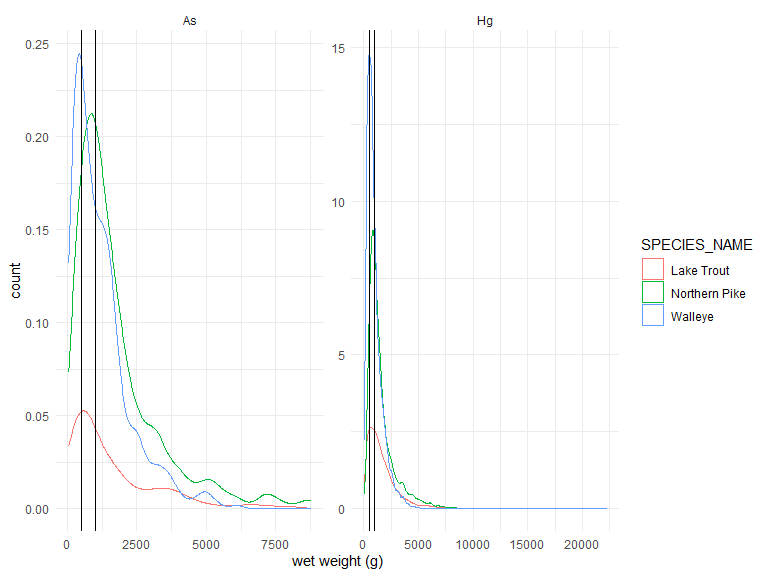


Figure 11.2: Number of fish with different contaminant concentrations. Vertical black lines show the weights that are commonly used for standardization.

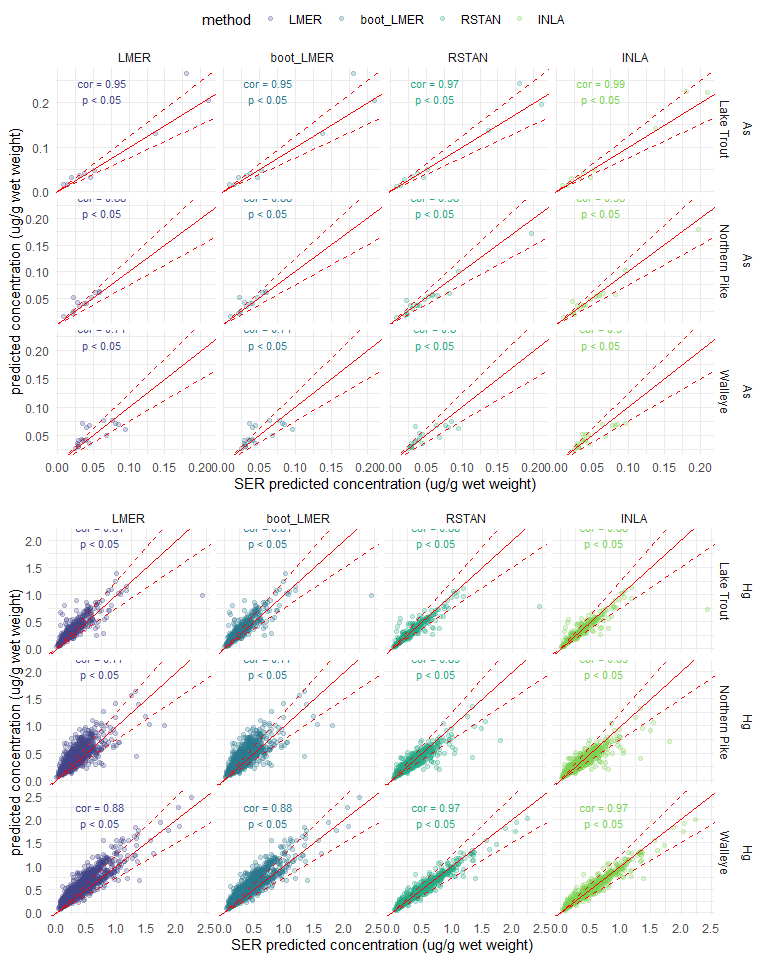


Figure 11.3: Predicted concentration of [As] and [Hg] in fish tissue of 500g fish compared to values predicted through the SER approach. Points display each individual fish prediction and a solid red line shows a 1-1 relationship. Dashed red lines provide an interval for predicted values that are within 25% of the predicted SER value. Correlation statistics are included in the top left of each panel.

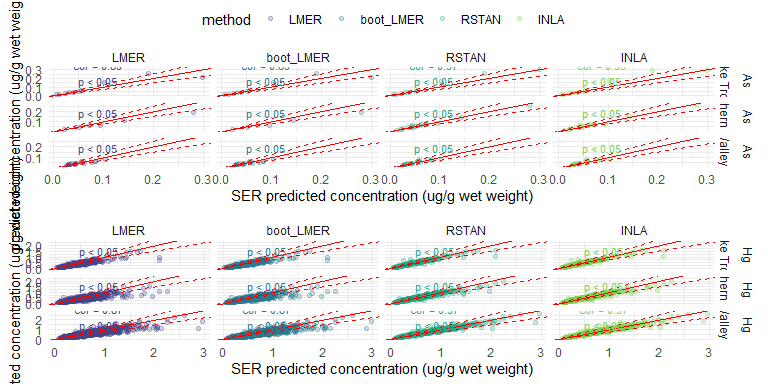


Figure 11.4: Predicted concentration of [As] and [Hg] in fish tissue of 1000g fish compared to values predicted through the SER approach. Points display each individual fish prediction and a solid red line shows a 1-1 relationship. Dashed red lines provide an interval for predicted values that are within 25% of the predicted SER value. Correlation statistics are included in the top left of each panel.

# References

(1) Canada, G. of. Mercury in Fish - Questions and Answers. *Government of Canada*, 2019.

(2) Ministry of Environment Conservation and Parks. Guide to Eating Ontario Fish. *ontario.ca*, 2023.

(3) Bhavsar, S. P.; Awad, E.; Mahon, C. G.; Petro, S. Great Lakes Fish Consumption Advisories: Is Mercury a Concern? *Ecotoxicology* **2011**, *20* (7), 1588–1598. <https://doi.org/10.1007/s10646-011-0731-0>.

(4) Kidd, K. A.; Muir, D. C. G.; Evans, M. S.; Wang, X.; Whittle, M.; Whittle, M.; Swanson, H. K.; Johnston, T. A.; Guildford, S. J. Biomagnification of Mercury Through Lake Trout (Salvelinus Namaycush) Food Webs of Lakes with Different Physical, Chemical and Biological Characteristics. *Science of The Total Environment* **2012**, *438*, 135–143. <https://doi.org/10.1016/j.scitotenv.2012.08.057>.

(5) Lescord, G. L.; Johnston, T. A.; Johnston, T. A.; Heerschap, M. J.; Keller, W.; Southee, F. M.; F. Meg Southee; O’Connor, C. M.; Constance M. O’Connor; O’Connor, C. M.; Dyer, R. D.; Branfireun, B. A.; Gunn, J. M. Arsenic, Chromium, and Other Elements of Concern in Fish from Remote Boreal Lakes and Rivers: Drivers of Variation and Implications for Subsistence Consumption. *Environmental Pollution* **2020**, *259*, 113878. <https://doi.org/10.1016/j.envpol.2019.113878>.

(6) Sumner, A. W.; Johnston, T. A.; Johnston, T. A.; Gretchen L. Lescord; Lescord, G. L.; Branfireun, B. A.; Gunn, J. M. Mercury Bioaccumulation in Lacustrine Fish Populations Along a Climatic Gradient in Northern Ontario, Canada. *Ecosystems* **2020**, *23* (6), 1206–1226. <https://doi.org/10.1007/s10021-019-00464-9>.

(7) Peng, X.; Liu, F.; Wang, W.-X. Organ-Specific Accumulation, Transportation, and Elimination of Methylmercury and Inorganic Mercury in a Low Hg Accumulating Fish. *Environmental Toxicology and Chemistry* **2016**, *35* (8), 2074–2083. <https://doi.org/10.1002/etc.3363>.

(8) Johnston, T. A.; Lescord, G. L.; Quesnel, M.; Savage, P.-L.; Gunn, J. M.; Kidd, K. A. Age, Body Size, Growth and Dietary Habits: What Are the Key Factors Driving Individual Variability in Mercury of Lacustrine Fishes in Northern Temperate Lakes? *Environmental Research* **2022**, 113740–113740. <https://doi.org/10.1016/j.envres.2022.113740>.

(9) Al-Sulaiti, M. M.; Soubra, L.; Al-Ghouti, M. A. The Causes and Effects of Mercury and Methylmercury Contamination in the Marine Environment: A Review. *Current pollution reports* **2022**, *8* (3), 249–272. <https://doi.org/10.1007/s40726-022-00226-7>.

(10) Eagles-Smith, C. A.; Ackerman, J. T.; Willacker, J. J.; Tate, M. T.; Lutz, M. A.; Fleck, J. A.; Stewart, A. R.; Wiener, J. G.; James G. Wiener; Evers, D. C.; Lepak, J. M.; Davis, J. A.; Pritz, C. F. Spatial and Temporal Patterns of Mercury Concentrations in Freshwater Fish Across the Western United States and Canada. *Science of The Total Environment* **2016**, *568*, 1171–1184. <https://doi.org/10.1016/j.scitotenv.2016.03.229>.

(11) Eckley, C.; Eagles-Smith, C.; Luxton, T.; Hoffman, J.; Janssen, S. E. Using Mercury Stable Isotope Fractionation to Identify the Contribution of Historical Mercury Mining Sources Present in Downstream Water, Sediment and Fish. *Frontiers in Environmental Chemistry* **2023**. <https://doi.org/10.3389/fenvc.2023.1096199>.

(12) Negrazis, L.; Kidd, K. A.; Erdozain, M.; Emilson, E. J. S.; Mitchell, C. P. J.; Gray, M. A. Effects of Forest Management on Mercury Bioaccumulation and Biomagnification Along the River Continuum. *Environmental Pollution* **2022**, 119810–119810. <https://doi.org/10.1016/j.envpol.2022.119810>.

(13) Kluke, C.; Lescord, G.; Johnston, T. A.; Kielstra, B.; Lock, A.; Bhavsar, S. P.; Gunn, J. Spatial Patterns and Environmental Factors Related to Arsenic Bioaccumulation in Boreal Freshwater Fish. *Canadian Journal of Fisheries and Aquatic Sciences* **2022**. <https://doi.org/10.1139/cjfas-2022-0106>.

(14) Weber, M. J.; Cashatt, D. Use of Occupancy Modelling to Identify Sample Sizes and Waterbodies with Fish Exceeding Mercury Consumption Advisory Thresholds. *Ecotoxicology* **2021**, *30* (2), 323–330. <https://doi.org/10.1007/s10646-020-02326-w>.

(15) St-Onge, I.; Magnan, P. Impact of Logging and Natural Fires on Fish Communities of Laurentian Shield Lakes. **2000**, *57*.

(16) Bhavsar, S. P.; Gewurtz, S. B.; McGoldrick, D. J.; Keir, M. J.; Backus, S. Changes in Mercury Levels in Great Lakes Fish Between 1970s and 2007. *Environmental Science & Technology* **2010**, *44* (9), 3273–3279. <https://doi.org/10.1021/es903874x>.

(17) Depew, D. C.; Burgess, N. M.; Robin, A. M.; Baker, R.; Bhavsar, S. P.; Bodaly, R. A. (Drew).; Eckley, C. S.; Evans, S.; Gantner, N.; Graydon, J. A.; Jacobs, K.; LeBlanc, J. E.; St. Louis, V. L.; Campbell, L. M. An Overview of Mercury Concentrations in Freshwater Fish Species: A National Fish Mercury Dataset for Canada. *Canadian Journal of Fisheries and Aquatic Sciences* **2013**, *70* (3), 436–451. <https://doi.org/10.1139/cjfas-2012-0338>.

(18) Fong, Y.; Rue, H.; Wakefield, J. Bayesian Inference for Generalized Linear Mixed Models. *Biostatistics* **2010**, *11* (3), 397–412. <https://doi.org/10.1093/biostatistics/kxp053>.

(19) Gabry, J.; Goodrich, B. [*Rstanarm: Bayesian Applied Regression Modeling via Stan*](https://CRAN.R-project.org/package=rstanarm); 2023.

(20) Rue, H.; Lindgren, F.; Teixeira Krainski, E. *INLA: Full Bayesian Analysis of Latent Gaussian Models Using Integrated Nested Laplace Approximations*; 2023.

(21) Rue, H.; Riebler, A. I.; Sørbye, S. H.; Illian, J. B.; Simpson, D. P.; Lindgren, F. K. [Bayesian Computing with INLA: A Review](http://arxiv.org/abs/1604.00860). *Annual Reviews of Statistics and Its Applications* **2017**, *4* (March), 395–421.

(22) Lescord, G. L.; Johnston, T. A.; Johnston, T. A.; Branfireun, B. A.; Gunn, J. M. Mercury Bioaccumulation in Relation to Changing Physicochemical and Ecological Factors Across a Large and Undisturbed Boreal Watershed. *Canadian Journal of Fisheries and Aquatic Sciences* **2019**, *76* (12), 2165–2175. <https://doi.org/10.1139/cjfas-2018-0465>.

(23) R Core Team. [*R: A Language and Environment for Statistical Computing*](https://www.R-project.org/); R Foundation for Statistical Computing: Vienna, Austria, 2023.

(24) Wickham, H. [*Tidyverse: Easily Install and Load the Tidyverse*](https://CRAN.R-project.org/package=tidyverse); 2023.

(25) Wickham, H.; Chang, W.; Henry, L.; Pedersen, T. L.; Takahashi, K.; Wilke, C.; Woo, K.; Yutani, H.; Dunnington, D. [*Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*](https://CRAN.R-project.org/package=ggplot2); 2023.

(26) Kassambara, A. [*Ggpubr: Ggplot2 Based Publication Ready Plots*](https://rpkgs.datanovia.com/ggpubr/); 2023.

(27) Bates, D.; Maechler, M.; Bolker, B.; Walker, S. [*Lme4: Linear Mixed-Effects Models Using Eigen and S4*](https://github.com/lme4/lme4/); 2023.