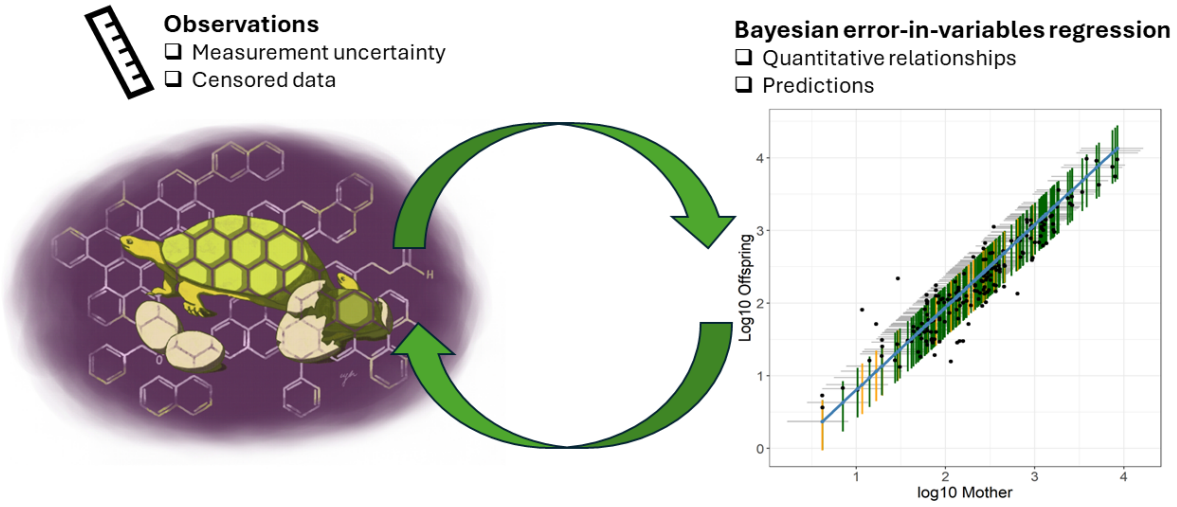
Quantifying the relationship between observed variables that contain censored values using Bayesian error-in-variables regression

Peter Vermeiren, Sandrine Charles, Cynthia C. Muñoz

Quantifying the relationship between observed variables that contain censored values using Bayesian error-in-variables regression

Peter Vermeiren[1,*], Sandrine Charles[2], Cynthia C. Muñoz[1]

[1]University of South-Eastern Norway, Dept. Natural Science and Environmental Health, Gullbringvegen 36, 3800 Bø, Norway
[2]Universite Claude Bernard Lyon 1; CNRS, UMR 5558, Laboratory of Biometry and Evolutionary Biology, 69100, Villeurbanne, France
[*]corresponding author peter.vermeiren@gmail.com



Abstract

We aimed to address two common challenges for scientists working with observational data: "how to quantify the relation between two observed variables", and, "how to account for censored observations" (i.e., observations whose value is only known to fall within a range). Quantifying the relationship between observed variables, and predicting one measured quantity from another (and vice versa), violates the assumption of standard regression regarding the existence of an independent, explanatory variable that is measured with no (or limited) measurement error. To overcome this challenge, we developed and tested a Bayesian error-in-variables, EIV, regression model which accounts for measurement uncertainty orthogonally. Moreover, parameter estimation using Bayesian inference allowed the full parameter uncertainty to be propagated into probabilistic model predictions suitable for decision making. Alternative model formulations were applied to a dataset containing measured concentrations of organic pollutants in mothers and their eggs from the freshwater turtle *Malaclemys terrapin* and validated against an independent dataset of the turtle *Chelydra serpentina*. The best performing EIV model was then applied to the dataset again after censoring observations in one or both variables. The Bayesian implementation allowed for such application as independent likelihoods for both censored and uncensored data could be combined. The EIV model performed well, as revealed by posterior predictive checks around 85%, and obtained comparable parameter estimates in both censored and uncensored cases. The resulting model allows scientists and decision-makers to quantitatively link measured variables, and make predictions from one variable to the next while accounting for measurement uncertainties and censored data.

keywords: Orthogonal regression, Measurement error, Maternal transfer, Reptile ecotoxicology, environmental pollution

1 Introduction

In ecology and ecotoxicology, researchers are often confronted with relationships between two observed variables. For example, allometric relations between various length and weight measurements of organisms [Vermeiren et al., 2021], measured concentrations of chemicals in different environmental matrices [Muñoz et al., 2024], relationships between different observed toxic effects, or when comparing the performance of two indices or measuring devices [Hawkins and Weckwerth, 2016, Connors et al., 2022]. The strength and direction of the association between such observed variables can be described in a correlation. However, capturing this relation quantitatively and predicting one measured quantity from another (and vice versa) faces a challenge. Regression is frequently used to describe the relationship between two or more variables, which can be linear or non-linear and amenable to several quantitative data types (e.g., counts, binary, or continuous data) commonly encountered in ecology and ecotoxicology. Nevertheless, when the variables are both observed features, they violate a basic assumption of standard regression models, namely the existence of an independent explanatory variable that is measured without uncertainty (i.e., error) and used to explain a dependent variable [Pallavi et al., 2022]. Often, measurement uncertainty is neglected in at least one of the two variables (which is then assumed to be the "independent" variable) leading to biased parameter estimates [Mikkonen et al., 2019]. Indeed, interchanging the role of the two variables as "dependent" and "independent" variables leads to different regression coefficients [Andreon and Hurn, 2013]. Hence, there is a need for a model that can mathematically quantify the relationship between the two variables, where each variable contains observations associated with measurement uncertainty.

Error-in-variables (EIV) regression methods have been proposed to tackle this challenge. They consider observations of each variable to come from some unobserved, true value with added measurement error [Mikkonen et al., 2019]. Orthogonal regression is one of the main EIV techniques [Pallavi et al., 2022] that aims to minimise the distance between observed data and the fitted regression curve orthogonally (i.e., perpendicular to the regression line) during parameter inference. In contrast, a standard regression aims to minimise the distance between observed data and the regression curve only in the direction of the dependent variable (that is usually vertically).

Capturing parameter uncertainties during inference is an important requisite for models to be useful for decision-making, for instance, in the context of risk assessment or environmental management [Schuwirth et al., 2019]. Here, a Bayesian approach toward parameter inference is well suited. A Bayesian approach considers parameters as random variables (rather than fixed but unknown point estimates under a frequentist approach) for which prior probability distributions are updated from the sample of the observations during the inference process [Gelman et al., 2014]. Consequently, parameter uncertainty (as shape, height, and width of the probability distributions) is explicitly quantified during inference, and can be propagated into predictions and in any decision-making process built around these predictions. Within a Bayesian context, the term Bayesian EIV regression has been more commonly used [Reily and Patino-Leal, 1981, Mikkonen et al., 2019, Splett et al., 2019], although the term orthogonal regression has also been used [Novick et al., 2012]. Irrespective of the terminology, and despite the advantage of allowing for explicit modelling of uncertainties in both measured variables under a Bayesian framework [Splett et al., 2019], the application of Bayesian EIV models is still relatively little

evaluated, as orthogonal regression has traditionally mainly relied on a frequentist framework for parameter inference. One particular issue that has been highlighted (under both frequentist and Bayesian frameworks) relates to the identifiability of model parameters (i.e., the theoretic possibility to learn the true values of the model's underlying parameters if an infinite set of observations was available, Splett et al. [2019]).

A further challenge in working with observed values is the possibility for censored data to occur (also referred to as truncated or non-detect data, Tobin [1958], Helsel [2005]). This is a situation in which an observation could not be determined to a specific value, but rather falls below, above, or between given bounds, thus called left, right, or interval-censored data, respectively. This is a relatively common occurrence in ecological and ecotoxicological data [Shoari and Dubé, 2018, Landes et al., 2020], for instance, when chemical concentrations fall below analytical detection limits, organisms survive or do not show an observable effect until some time after the study has ended, or events occur during gaps in observations. The inclusion of censored data in linear regression is generally known as Tobit regression, yet, the combination of EIV with censored data is much less established [Wang, 1998, Hawkins and Weckwerth, 2016]. The inclusion of censored data during inference of model parameters is generally achieved by splitting the likelihood (i.e., the probability of the data given the parameters) into a product of individual likelihoods for the censored and uncensored data. Such a product of likelihoods is relatively straightforward to implement under a Bayesian framework through the use of Bayes theorem [Kon Kam King et al., 2015, Hawkins and Weckwerth, 2016, Qi et al., 2022].

To tackle the challenge of quantifying the relationship between observed variables in ecology and ecotoxicology, we developed a Bayesian EIV model that is parsimonious regarding model parameters; that accounts for uncertainties in both variables; that propagates these uncertainties into predictions relevant to decision-making; and that accounts for censored data. To illustrate our strategy, several variants of the EIV model have been applied to a dataset containing measured concentrations of organic pollutants in mothers and their eggs from the freshwater turtle *Malaclemys terrapin* and validated against an independent dataset of the turtle *Chelydra serpentina*. The best performing EIV model was then applied to the dataset again after censoring observations in one or both variables. This case study is detailed below after all features of the model have been described.

2 Methods

2.1 Model concept and formulation

We developed a Bayesian EIV model (equation. 1) inspired by Novick et al. [2012]. Specifically, for each pair of n independent observations, X_i^{obs} and Y_i^{obs} $(i = 1...n)$ represent the observed values of the two variables. The values of both X_i^{obs} and Y_i^{obs} are observed with uncertainty, and are linked to their respective latent (unobserved) values X_i and Y_i, which represent the "true" value of the variable without uncertainty. This stochastic link is formalised as a normal distribution with mean value at X_i (resp. Y_i) and standard deviation σ_u^2 (resp. σ_y^2) for X_i^{obs} (resp. Y_i^{obs}). The assumption that observed data stem from a distribution around the true, latent value is also called a classical measurement error model [Andreon and Hurn, 2013]. In addition, latent values X_i are assumed to be independent, identically distributed random variables stemming from a normal distribution (i.e., a structural model according to Novick et al. [2012]) around population mean μ_x

with standard deviation σ_x^2; such an assumption prevents extreme values during model calibration. Finally, the deterministic link between the two latent variables is characterised by intercept, α, and slope, β, parameters which characterise the linear dependency. The following set of equations describes the base Bayesian EIV model (further: **model M6**) with six parameters gathered in vector $\theta_6 = \{\alpha, \beta, \sigma_u^2, \sigma_y^2, \mu_x, \sigma_x^2\}$.

$$\begin{aligned} X_i^{obs} &\sim \mathcal{N}(X_i, \sigma_u^2) \\ Y_i^{obs} &\sim \mathcal{N}(Y_i, \sigma_y^2) \\ X_i &\sim \mathcal{N}(\mu_x, \sigma_x^2) \\ Y_i &= \alpha + \beta X_i \end{aligned} \tag{1}$$

In practice, since we assume that X_i^{obs} is randomly distributed, the deterministic parameterization of Y_i can equivalently be derived from X_i^{obs} instead of X_i (equation. 2).

$$Y_i = \alpha + \beta X_i^{obs} \tag{2}$$

Alternative formulations were developed to improve model performance and parsimony (i.e., reduce the number of required parameters, fig. 1), and considering previous reports on identifiability issues in EIV models [Splett et al., 2019]. A first alternative (**model M5**) assumes that the uncertainty on both observed variables, X_i^{obs} and Y_i^{obs}, is equal (e.g., when both are measurements of the same process collected with the same methodology). This reduces the number of parameters to 5, that is $\theta_5 = \{\alpha, \beta, \sigma^2, \mu_x, \sigma_x^2\}$. A second alternative (**model M4**) additionally assumes that the slope β has a fixed value of 1 which corresponds to a relationship between X and Y that does not vary with increasing or decreasing values. In this case, the linear dependency gets reduced to $Y_i = \alpha + X_i$. Model M4 is thus described by 4 parameters $\theta_4 = \{\alpha, \sigma^2, \mu_x, \sigma_x^2\}$.

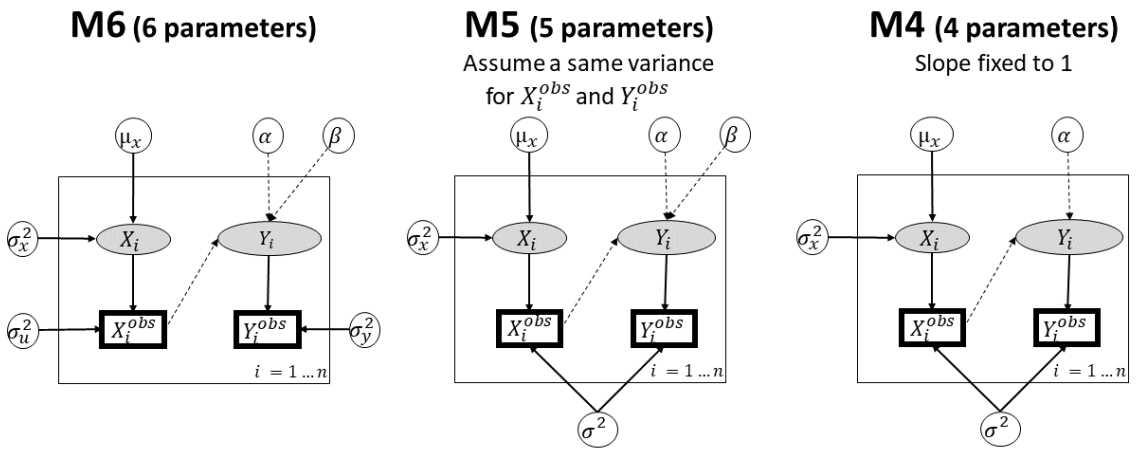


Figure 1: Schematic representation of model alternatives as directed acyclic graphs; open circles: parameters, filled ellipses: variables, squares: observations, full arrows: stochastic links, dashed arrows: deterministic links.

The model was adapted to account for censored data. Therefore, the likelihoods (i.e., the probability of the observed data given the parameters) were calculated independently for uncensored and censored data, and then combined using Bayes theorem into the product of these individual likelihoods. In practice, this means that each data point, Y_i^{obs}, that was censored (further referred to as Y_i^{cens}), is no longer a continuous variable, but rather a binary outcome Z_i (censored or not). This outcome was represented as a Bernoulli distribution of probability p_i, itself determined as the cumulative probability for the censored data to be below or above the censoring limit named *cut*, the censored data being itself sampled from a normal probability distribution around mean Y_i and standard deviation σ_y^2 (equation. 3).

$$\begin{aligned} Z_i &\sim Bern(p_i) \\ p_i &= F(Y_i^{cens}, cut) \end{aligned} \tag{3}$$

with F the cumulative distribution function of $\mathcal{N}(Y_i, \sigma_y^2)$. In case of left-censored data, $F(Y_i^{cens}, cut) = P(Y_i^{cens} \leq cut)$.

2.2 Modelling strategy

Two analyses were conducted to develop and apply the model to datasets containing censored data. In the first analysis, we aimed to develop and test the Bayesian EIV model using two datasets. The first dataset was used to calibrate the three model alternatives (M6, M5, and M4, section 2.4), to select the best performing alternative (section 2.5), and to internally validate the chosen alternative (Fig. 2, section 2.6). Therefore, we used a

3-fold cross-validation where the dataset was randomly split into three equal parts. Then, two thirds (i.e., two subsets of the data) were combined into one training set, and the remaining third was used as the corresponding test set. This partitioning was done three times, resulting in three different training and test set combinations. Calibration of the three model alternatives and selection of the best one were then done on the three training sets, while the internal validation was done on the complementary three test datasets. This three-fold cross-validation was chosen to evaluate the influence of the input datasets on the models' performance, while keeping a sufficiently large number of data points within each fold. A second, independent, dataset was subsequently used for external validation of the best performing model, with the parameter values for this model taken across the three-fold calibration.

In a second analysis, the best performing model was tested for its ability to account for censored data. Therefore, the model was re-calibrated to the whole first dataset, but with part of this dataset censored. Specifically, the lower 20^{th} percentile of the data corresponding to Y_i^{obs} was censored (i.e., recorded as a value falling below the 20^{th} percentile). It was then assessed (1) how much parameter distributions differed between the calibration against the uncensored and the censored dataset, and how well the model represented the original values (before censoring) during validation. This procedure was repeated, but then censoring the lower 20^{th} percentile of the data corresponding to X_i^{obs}, and once more censoring the lower 20^{th} percentile of the data in both X_i^{obs} and Y_i^{obs} (Supplement S1).

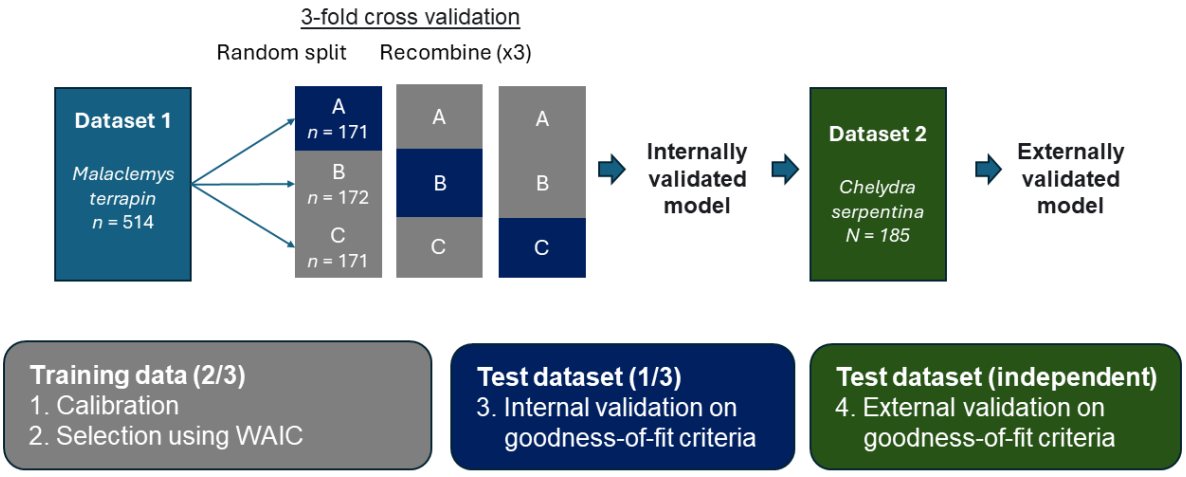


Figure 2: Schematic representation of our modelling strategy.

2.3 Implementation

Parameter estimation was conducted using Bayesian inference to allow the full parameter uncertainty to be quantified and propagated into probabilistic model predictions. Parameter estimation was performed using Monte Carlo Markov Chain (MCMC) simulations with Gibbs sampling using the JAGS software via the R2jags package version 0.7-1 [Yu-

Sung and Masanao, 2021], in R version 4.3.2 [R Core Team, 2023], on three independent chains. JAGS model code is presented in Supplement S1.

2.4 Calibration

Preliminary model runs using Raftery and Lewis' diagnostic were used to assess suitable lengths for the MCMC burn-in and sampling phases [Raftery and Lewis, 1992]. This diagnostic estimates the number of MCMC iterations needed to obtain a sufficiently precise posterior (specifically to estimate the 0.025 quantiles of the posterior distribution with an accuracy of 0.005). The values suggested by Raftery and Lewis' diagnostic were rounded to the nearest 10 and tripled to account for high autocorrelation. At least 100 burn-in iterations were used, and chains were thinned by 10 for computational efficiency (Supplement S2).

MCMC chain convergence was assessed as the overlap of the independent chains on MCMC traceplots and via the Rhat diagnostic (using <1.01 as the criterion for convergence of individual parameters, Gelman and Rubin [1992]). Autocorrelation was checked on traceplots and using the effective sample size. Parameter prior and posterior probability distributions were compared, and bi-variate scatterplots of posterior parameter distributions were checked for collinearity in parameter estimates.

We opted to define broad, vague prior parameter distributions. This ensures parameter inference to be driven largely by the information within the data. These priors were derived based on logical reasoning about parameter ranges, and information about the maximum observed value across the whole dataset (a data point which was then removed from further analyses). Priors on σ_u^2 and σ_y^2 (or simply σ^2 in models **M5** and **M4**), and on the population level standard deviation σ_x^2, were given on their respective precision τ_u, τ_y, (or simply τ) and τ_x where the precision equals the inverse of the variance. Gamma priors were used for the precision, with a shape coefficient of 1 and a rate of 0.001. Normal distributions were used to set the prior for parameters: β (the slope of the line) with a mean of 1 and a standard deviation of 10; α (the intercept of the line) with a mean of 0 and the maximum observed value in the data as the standard deviation; and μ_x (the population mean) with the half of the maximum observed value as the mean, and the maximum observed value as the standard deviation.

2.5 Model selection

To compare model alternatives (**M6**, **M5**, and **M4**), we evaluated the predictive performance of each model using 3-fold cross-validation. For each model alternative, and for each training set, the within-sample predictive performance was assessed using the Widely Applicable Information Criterion (WAIC, Watanabe [2013]). This is a highly suitable information criterion in a Bayesian context since it considers the full uncertainty in the model by using the full posterior density for testing predictive capacity, as compared to parameter point estimates as done in AIC and Deviance Information Criterion, DIC [Gelman et al., 2014]. Plots of residuals for all model alternatives were also checked for any remaining pattern.

2.6 Validation

After selection, the best performing model was validated internally and externally. For internal validation, the out-of-sample predictive accuracy of the model was assessed by using the model calibrated on each of the three training datasets to predict observations in their three respective test datasets. For external validation, predictions were made for a separate, independent dataset (2^{nd} dataset) with the full calibrated model (using the posterior parameter estimates across all training datasets). We focused on the model's ability to predict Y_i^{obs} from X_i^{obs} (in line with the traditional use of regression to predict one variable from another, even though in EIV models the role of both variables is interchangeable).

For internal and external validation, several goodness-of-fit measures were assessed, namely: (i) the posterior predictive check (PPC) which was calculated and plotted to identify which percentage of observed Y_i^{obs} values fell within the 95% credible interval of model predictions (with values closer to 95% indicating better performance) ; (ii) Pearson correlation coefficients for prediction (r^2) to evaluate how closely observed and predicted Y_i^{obs} values align (values closer to 1 indicating better alignment); (iii) the Nash-Sutcliffe coefficient of Efficiency (NSE) to compare the relative magnitude of residual against measured data variance (values closer to 1 indicating better fit); and (iv) the normalized Root Mean Squared Error of prediction (nRMSE) to assess the overall magnitude of residual variance in comparison to the magnitude of the mean (with values of 0.5 or lower preferred as they indicate a variance less or equal to half of the mean observed value).

Regarding the application of the best performing model towards censored datasets, the same four goodness-of-fit criteria (PPC, r^2, NSE, and nRMSE) were calculated by comparing the predicted Y_i^{obs} values (after re-calibrating the model to the censored first dataset) against the uncensored data. Meanwhile, for the censored data, it was assessed how many were predicted below the detection limit, and how well they correlated with the original values before censoring.

2.7 Model application

We applied model alternatives **M6**, **M5**, and **M4** in a case study of the maternal transfer of organic pollutants between mothers and their eggs. Such transfer represents an exposure route for developing embryos during a sensitive developmental stage and before encountering the external environment (Hitchcock et al. [2017], Muñoz and Vermeiren [2018], Gómez-Roig et al. [2021]). This can affect embryo vitality and survival, as well as health and disease later in life (Hamlin and Guillette Jr [2011], Chin et al. [2013], Kim et al. [2018], Basak et al. [2020]).

A large database containing the current state-of-the-art data regarding concentrations of organic pollutants measured in paired samples of reptile mothers and their eggs was recently collected through a systematic search for published data, followed by their extraction, homogenisation, and integration into one comprehensive database [Muñoz et al., 2024]. All pollutant concentrations in this database were standardized to $ng.g^{-1}$ lipid basis. We selected two subsets of this large database to apply the models. Specifically, we selected a homogeneous dataset for calibration, selection, and internal validation (dataset 1). This dataset contained measurements of polychlorinated biphenyls (PCBs), organochlorine pesticides (OCPs) and polybrominated diphenyl ethers (PBDEs) measured in female livers and their whole egg in the freshwater turtle *Malaclemys terrapin*, originally

collected by Basile et al. [2011]. The dataset contained some measurements below the detection limit, however, with no further information regarding the detection limit reported in the original study. Consequently, only values above detection limits (i.e., uncensored data) were used, resulting in a database with 514 observations. Additionally, for external validation, we selected a comparable dataset, which contained a set of 14 PCBs and OCPs that were also recorded in dataset 1 and were measured on the same tissues (mothers' liver and offsprings' whole egg); however, these measurements were obtained from a different species, namely the freshwater turtle *Chelydra serpentina*, in a study conducted by Hebert et al. [1993]. To link the datasets to the model, log_{10} lipid-normalized pollutant concentrations observed in mothers' liver were used as X_i^{obs} values, while observations in offsprings' eggs were used as Y_i^{obs} values.

3 Results

3.1 Selection

The formulation of model M6 led to identifiability issues. Specifically, the MCMC algorithm was highly autocorrelated leading to low effective sample sizes of less than 100 for α and β, and had convergence issues with Rhat values close to 1.1 for cross-validation run 2 (Supplement S3). Moreover, parameter values for standard deviations σ_u^2 and σ_y^2 (estimated as their respective precision τ_u and τ_y) on X_i^{obs} and Y_i^{obs} were pushed towards their lower limit; and bimodal marginal posterior distributions were returned for the intercept α and the slope β of the regression line (Supplement S4). Given these issues, model M6 was not further considered.

Meanwhile, models M5 and M4 converged well, with clearly identified, narrow marginal posterior parameter distributions distinct from the prior ones (Table 1, Supplement S4). Marginal posterior parameter distributions were highly comparable across the three training datasets, with most variation in the precision τ (Fig. 3). Parameters in model M5 and M4 were clearly uncorrelated, with exception of parameters α and β in model M5 ($corr = -0.873$, Fig. 3, Supplement S4). Comparing both models, M5 performed better, as revealed by a lower mean ($\pm$ standard deviation) WAIC of 6.37 ± 83.03 across three training sets compared to 57.83 ± 75.05 for model M4. Model M5 was therefore selected as the best performing alternative and used for further evaluation.

	Prior		Posterior		
	median	95% CI	median	95% CI	median ÷ 95% CI
α	0.00	[-5.74 ; 5.74]	-0.33	[-0.41 ; -0.25]	2.13
β	1.02	[-18.53 ; 20.56]	1.14	[1.08 ; 1.19]	10.13
τ	695.24	[25.26 ; 3684.00]	24.91	[19.05 ; 30.27]	2.22
μ_x	1.46	[-4.31 ; 7.21]	1.22	[1.14 ; 1.29]	7.97
τ_x	695.24	[25.26 ; 3684.00]	2.70	[2.23 ; 3.27]	2.60

Table 1: Prior and posterior parameter quantiles in model M5; median, 95% confidence interval (CI), and ratio of median over 95% CI.

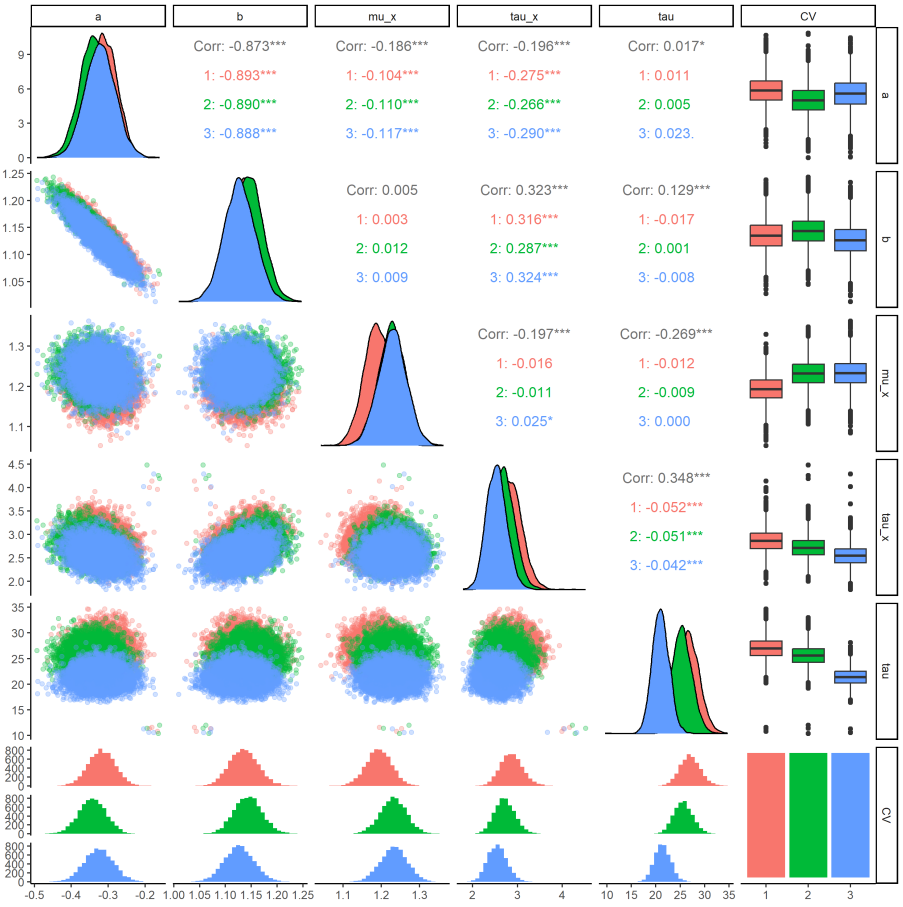


Figure 3: Posterior parameter correlations in model M5 coloured per training set.

3.2 Validation

Model M5 performed well during internal and external validation, as revealed by the goodness-of-fit criteria. Specifically, for internal validation: (i) the PPC ranged (min. - max.) between 82.5% and 90.1% of measured pollutant concentrations in eggs in the 3-fold test datasets to fall within the 95% credible interval of predictions (Fig. 4); (ii) the predictive Pearson coefficient r^2 ranged between 0.90 and 0.93 indicating close alignment between observed and median predicted values; (iii) the NSE ranged between 0.79 to 0.86 suggesting that variance is well captured by the model; and (iv) low nRSME between 0.25 and 0.31 were observed. Results of the external validation against an independent dataset of pollutant concentrations in the turtle *C. serpentina* were highly comparable to those of the internal validation with a PPC of 84.9, an r^2 of 0.93, an NSE of 0.84 and an nRMSE of 0.12.

3.3 Interpretation

The 95% credible interval for the slope parameter β in model M5 did not include 1 (Table 1), suggesting a trend for higher maternal transfer at higher mother concentrations. In model M5, the intercept α differed from zero, and the precision τ capturing the uncertainty around the observed variables X_i^{obs} and Y_i^{obs} was about 10 times larger (i.e., small standard deviation) compared to the overall population level precision, τ_x (Table 1).

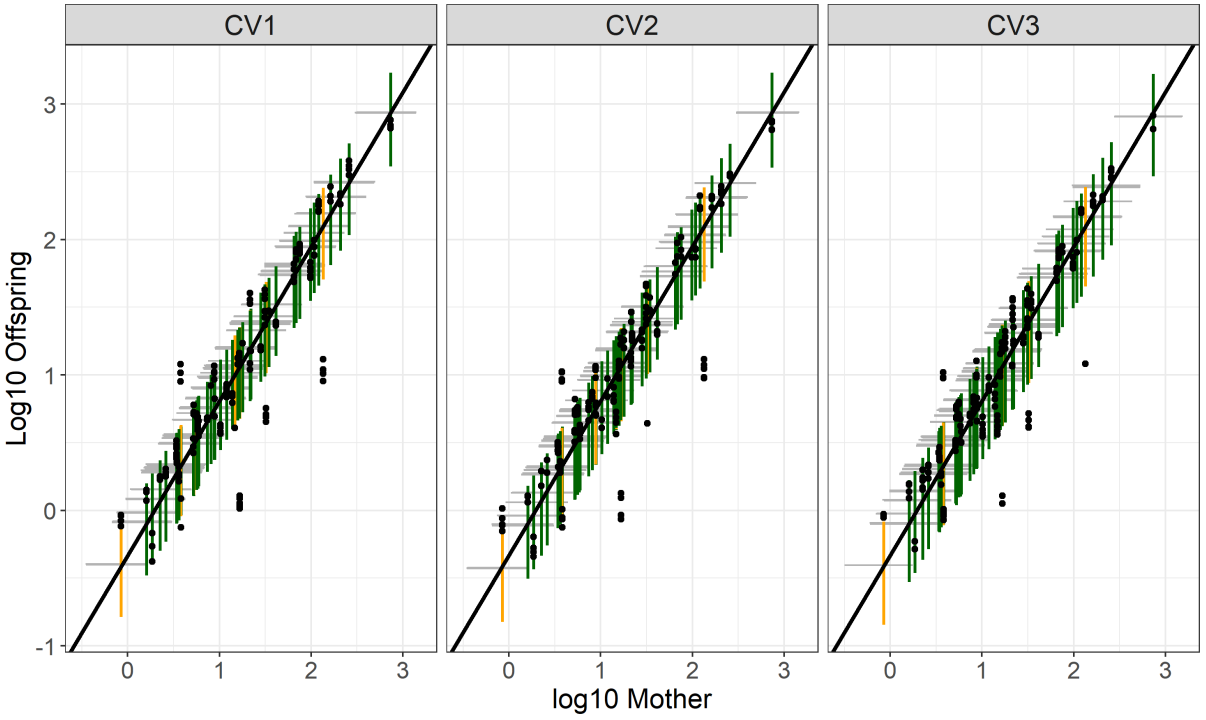


Figure 4: PPC plots for model M5 across the three cross validation (CV) test-folds in dataset 1; segments in green (resp. orange) indicate 95% credible intervals overlapping (resp. not-overlapping) with observed data.

3.4 Application to censored data

Applying model M5 to a dataset where the lowest 20% of the data was censored in one or both of the observed variables, resulted in a good model convergence as revealed by overlapping MCMC traceplots for the three chains showing no pattern, Rhat values for all parameters close to 1.00, and sufficient effective sample sizes (Supplement S3). Additionally, in all cases, narrow posterior probability distributions were obtained compared to the prior distributions, and overlapping those obtained when applying model M5 to the same dataset without any censored values (Fig. 5, Supplement S4). Consequently, the fitted regression lines highly aligned across results with the uncensored and censored data (Fig. 6).

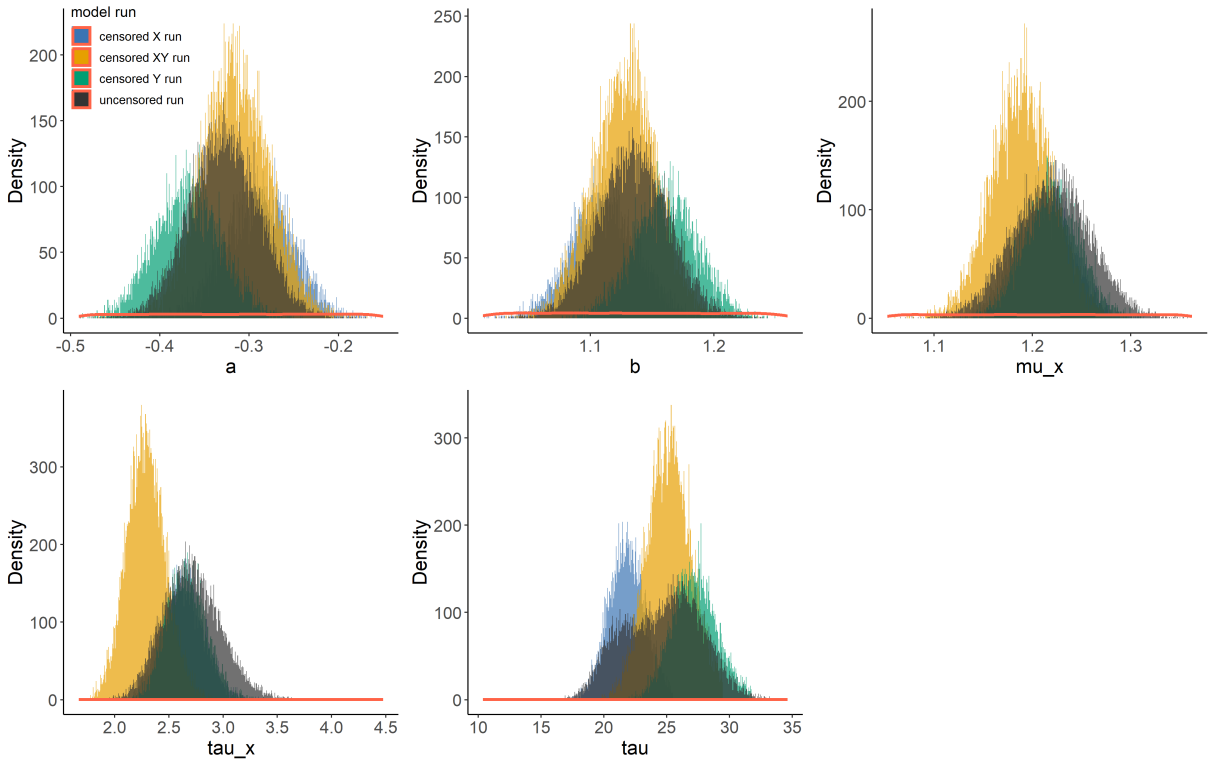


Figure 5: Probability density plots for M5 model parameters, including prior (red line) and posterior (histograms coloured based on whether model M5 was calibrated against dataset 1 with or without censored values)

As expected, the intercept α and the slope β were highly correlated (-0.926, -0.912, -0.921 when censoring only in Y_i^{obs}, only in X_i^{obs}, or in both variables, respectively), similarly to the results obtained for the uncensored dataset (fig. 11). Other parameters were uncorrelated.

Model M5 performed well when applied to the dataset censored only in Y_i^{obs}, only in X_i^{obs}, or in both variables, with a PPC of 85.64%, 88.49%, and 90.33%, respectively, of uncensored concentrations in eggs falling within the 95% credible interval of predictions. Additionally, for these uncensored data, good performance was revealed by Pearson r^2 values of 0.89, 0.89, and 0.91, NSE of 0.78, 0.78, and 0.82, and nRSME of 0.22, 0.25 and 0.20 for the dataset censored only in Y_i^{obs}, only in X_i^{obs}, or in both variables, respectively.

Meanwhile, for the dataset with censored Y_i^{obs}, median predicted values for censored data points fell largely below the censoring limit, although 12.6% were predicted above it. Additionally, the median predicted values correlated poorly to the original values (i.e., before censoring), with a Pearson r^2 of 0.20. Yet, a group of outlying observations with relatively high maternal concentrations, but low offspring concentrations (compared to the bulk of observations in the dataset) could contribute to this poor correlation. Omitting the outlying values led to much better agreement between observed and predicted values, with just 1% above the censoring limit and a correlation with the original uncensored values of 0.52. (Note that a similar analysis was not conducted for the datasets censored in X_i^{obs} only and in both variables together. Indeed, it is not sensible to evaluate the predictive performance for censored datapoints as there are no known X_i^{obs} values to predict from, only the value of the censoring limit.)

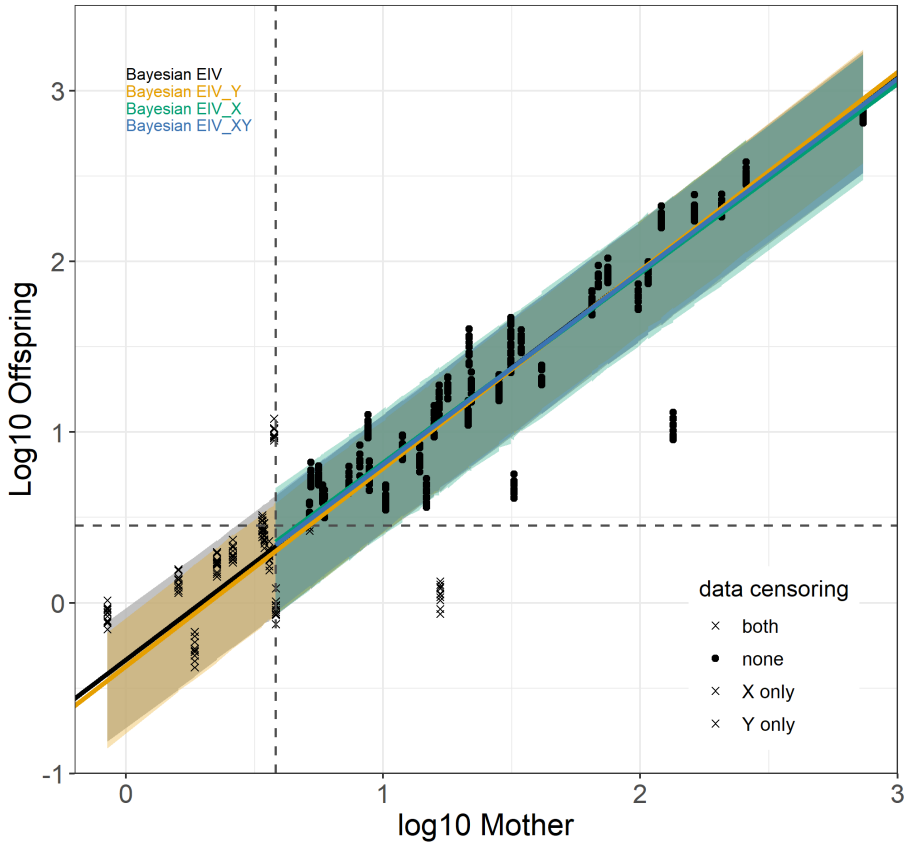


Figure 6: Median predictions and 95% credible interval band for model M5 on uncensored (black), Y censored (yellow), X censored (green, or both censored (blue) data. For datasets containing censoring in X, predictions below the detection limit use the detection limit as X value.

4 Discussion

We developed and tested a Bayesian EIV model allowing us to quantify the relationship between two observed variables, both associated with measurement uncertainty, and with the possibility to contain censored data. These are two challenges often encountered in ecology, ecotoxicology, and other scientific fields [Connors et al., 2022]. The EIV model allows to overcome a strong limitation in standard regression imposed by the assumption that uncertainty is only present in the dependent variable, while the explanatory variable is assumed to be measured with no (or very limited) uncertainty [Pallavi et al., 2022].

4.1 Model implementation and performance

Quantifying the uncertainty on both variables, however, represents an additional challenge. Indeed, our first model alternative, M6, which included all 6 parameters $\theta_6 = \{\alpha, \beta, \sigma_u^2, \sigma_y^2, \mu_x, \sigma_x^2\}$ faced identifiability issues as it attempted to push the variances of at least one of the two variables towards their lower limit (i.e., towards zero which is the minimum variance possible, as also determined by their gamma priors). Meanwhile, both slope and intercept parameters returned bimodal marginal posterior distributions. In other words, the model attempted to align with classical regressions of either "X versus

Y" or "Y versus X", but struggled to find a way to divide the uncertainty between the two variables. Parameter identifiability has been outlined as a challenge in EIV models [Splett et al., 2019]. Novick et al. [2012] were able to estimate the full 6-parameter model, since their data included repeated measurements of the same samples thus allowing the quantification of measurement uncertainty in each variable separately. Collecting data with repeated measurements so that measurement uncertainty can be explicitly quantified would thus be a valuable recommendation to improve study designs for statistical analyses if resources allow. Nevertheless, this recommendation does face challenges related to the time and effort needed to collect such data.

An alternative approach is to constrain the variances to a ratio (σ_u^2/σ_y^2), based on assumptions or selected measurements, thereby reducing the problem to a 5-parameter estimation [Novick et al., 2012, Andreon and Hurn, 2013]. Our implementation of M5 applies this alternative by assuming that the variance is equal between the two variables (i.e., a ratio of 1:1). Such assumptions, however, need to be clearly communicated, so that their applicability can be checked and potential effects on parameter estimates evaluated [Hawkins and Weckwerth, 2016]. Nevertheless, model assumptions or implementations are not always communicated when research is predominantly focused on the application and outcomes of the technique. For instance, orthogonal regression was applied to assess the equivalence of ecotoxicity tests between two *Daphnia* species [Connors et al., 2022], yet, the underlying model and assumptions were unspecified in the publication (and not upfront communicated in the documentation of the R package used to fit the model, Borchers [2023]). Increased details in the documentation would improve awareness and understanding of EIV models and their underlying assumptions. In our case study, both X_i^{obs} and Y_i^{obs} were measurements of pollutant concentrations in biological tissues. Hence, an assumption of equal variance is reasonable. We applied our model to data already available from the literature (which did not include repeated measurements of the same sample). Nevertheless, a recommendation for future studies would be to measure a few samples repeatedly, so that the assumption of equal variance can be tested, and, if needed, adjusted to a ratio between the two variances.

The 5-parameter model, M5, performed well in both internal and external validation, in compliance with our four goodness-of-fit criteria. This suggests that, given the assumption of equal variance on the measured variables, Bayesian orthogonal regression models can be used to quantify the linear relationship between measured variables. Moreover, the regression parameters were stable across the three training sets, although the variance, τ, on the observed variables did vary somewhat. The training datasets, each of which had 342 or 343 data points, were of sufficient size compared to rough rules of thumb of about 15 to 25 times more data than parameters in generalized linear regression models (e.g., Zuur et al. [2015]). In ecotoxicology, datasets can sometimes be quite small given ethical and practical constraints to experimentation [Muñoz et al., 2023]. Therefore, evaluating the behaviour of the model with lower numbers of data points would be a future research direction.

We applied the model to pollutant concentrations, which are quantitative continuous variables that are typically assumed to follow a Gaussian error distribution, as implemented in our equations (1). The model, however, could easily be extended towards count or ratio variables (as encountered in ecology and ecotoxicology, for instance, as numbers of offspring or proportions of survivors), by adapting to other error distributions (e.g., Poisson or Binomial distributions). The Bayesian implementation of the EIV model

is well suited to such changes, as it explicitly formulates the stochastic part of the model, which might be less visible in implementations under a frequentist framework.

4.2 Application to censored data

The Bayesian framework in which we developed the error-in-variables model lends itself well for the consideration of censored data as the likelihood in Bayes theorem can be split in independent likelihoods for censored and uncensored data [Qi et al., 2022]. We observed good performance for model M5 when applied to a dataset with the lower 20% of data in either one or both of the variables censored, with parameter estimates overlapping those for the same model calibrated to the uncensored dataset (fig. 6). This offers interesting opportunities for ecology and ecotoxicology where censored data are common. Particularly, it allows for the full inclusion of the information contained in the dataset into the model, rather than omitting censored data or substituting them with an arbitrary value thereby introducing bias and loss of information to the model results [Helsel, 2005, Hawkins and Weckwerth, 2016]. Moreover, the Bayesian implementation allows for explicit quantification of parameter uncertainties during inference, where censored EIV models thus far have relied on frequentist approaches where parameter confidence intervals were obtained as a secondary step after inference [Wang, 1998, Hawkins and Weckwerth, 2016]. Further research could consider the influence of right and interval censored data in addition to the left censored implementation presented here, and evaluate the influence of the amount of data and proportion of censored values within datasets on the stability and performance of the model fitting outputs.

4.3 Ecotoxicological interpretation

Our research focused on the development and testing of the Bayesian EIV model, with a detailed ecotoxicological analysis beyond the current scope (see e.g., Muñoz et al. [2024] for an investigation of data regarding maternal transfer of organic pollutants in reptiles). Nevertheless, we briefly discuss the parameterization of M5. Specifically, the intercept close to 0 and the slope close to 1 suggest that, overall, pollutant concentrations (on a lipid normalised basis) are comparable in offspring (whole egg) to that in mothers (liver). This implies, firstly, that maternal transfer of organic pollutants might pose a non-trivial exposure route for developing embryos of freshwater turtles such as *M. terrapin* and *C. serpentina* to the PCB, OCP and PBDE compounds which can affect embryo and juvenile development as demonstrated for several reptile species [Hamlin and Guillette Jr, 2011]. Secondly, for mothers, maternal transfer presents a potentially relevant pollution offloading route. Such offloading can decrease the concentration of pollutants in females and contribute to explain differences in body burdens of some pollutants between males and females [Humphries et al., 2021], although other processes also contribute to these differences such as differences in metabolic rates, lipid storage and allocation, and habitat usage as observed in wildlife (Binnington and Wania [2014], Lawson et al. [2020]) and humans [Salihovic et al., 2012]. Finally, for egg predators, this might lead to a source of dietary exposure [Warwick et al., 2013].

The fact that the estimated regression slope β was slightly larger than 1 (and the related intercept α slightly below 0), indicates that maternal transfer of organic pollutants is elevated at higher maternal concentrations. This result could hint at some underlying

biological mechanism which, for instance, limits maternal transfer at lower concentrations, or promotes higher transfers at elevated concentrations; increases (or decreases) elimination via other means in the mother (or egg) at elevated concentrations; or a combination of both. Alternatively, the difference in slope could be a result of difficulties in measuring lower concentrations compared to higher ones, although the good performance of the model against an external, independent dataset collected by different authors would make such a methodological reason less likely. At this moment, we can only hypothesise about the potential reasons. Nevertheless, further interpretation and exploration of the model's ecotoxicological outputs is needed. This includes testing the relevance and applicability of the model to other reptile species and other organic pollutants. Such further ecotoxicological work, using, among others, our implementation of Bayesian EIV, is an urgent research topic given the relatively long lifespan of many reptile species (and associated long-term accumulation and potential maternal transfer of pollutants); the limited knowledge on maternal transfer in reptiles; and the overall lag in ecotoxicology in this group compared to other vertebrates [EFSA PPR panel et al., 2018].

5 Acknowledgements

This research received funding through the EU Cost action "Periamar", the Graduate School "H2O'Lyon" (ANR17-EURE-0018), and the Erasmus+ program. We are grateful to Basile et al. 2011 and Hebert et al. 1993 for discussion and access to their data.

6 CRediT roles

All: Conceptualization; Funding acquisition; Methodology; **PV, CCM**: Investigation; Data curation; Formal analysis, Software; Validation; Visualization; **PV**: Writing - original draft; **SC, CCM**: Writing - review & editing.

References

S. Andreon and M. Hurn. Measurement errors and scaling relations in astrophysics: A review. *Statistical Analysis and Data Mining*, 6:15–33, 2013.

S. Basak, M. K. Das, and A. K. Duttaroy. Plastics derived endocrine-disrupting compounds and their effects on early development. *Birth defects research*, 112(17):1308–1325, 2020.

E. R. Basile, H. W. Avery, W. F. Bien, and J. M. Keller. Diamondback terrapins as indicator species of persistent organic pollutants: using barnegat bay, new jersey as a case study. *Chemosphere*, 82(1):137–144, 2011.

M. J. Binnington and F. Wania. Clarifying relationships between persistent organic pollutant concentrations and age in wildlife biomonitoring: individuals, cross-sections, and the roles of lifespan and sex. *Environmental toxicology and chemistry*, 33(6):1415–1426, 2014.

H. Borchers. *pracma: Practical Numerical Math Functions*, 2023. URL `https://CRAN.R-project.org/package=pracma`. R package version 2.4.4.

S. Y. Chin, J. D. Willson, D. A. Cristol, D. V. Drewett, and W. A. Hopkins. Altered behavior of neonatal northern watersnakes (nerodia sipedon) exposed to maternally transferred mercury. *Environmental Pollution*, 176:144–150, 2013.

K. A. Connors, J. L. Brill, T. Norberg-King, M. G. Barron, G. Carr, and S. E. Belanger. Daphnia magna and ceriodaphnia dubia have similar sensitivity in standard acute and chronic toxicity tests. *Environmental toxicology and chemistry*, 41:134–147, 2022.

EFSA PPR panel, C. Ockleford, P. Adriaanse, P. Berny, T. Brock, S. Duquesne, S. Grilli, A. Hernandez-Jerez, S. Bennekou, M. Klein, T. Kuhl, R. Laskowski, K. Machera, O. Pelkonen, S. Pieper, M. Stemmer, I. Sundh, I. Teodorovic, A. Tiktak, C. Topping, G. Wolterink, A. Aldrich, C. Berg, M. Ortiz-Santaliestra, S. Weir, F. Streissl, and S. R.H. Scientific opinion on the state of the science on pesticide risk assessment for amphibians and reptiles. *EFSA Journal*, 16:5125, 2018.

A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472, 1992.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis, 3rd edition*. Texts in statistical science. Taylor & Francis Group, 2014. ISBN 9781439840955.

M. D. Gómez-Roig, R. Pascal, M. J. Cahuana, O. García-Algar, G. Sebastiani, V. Andreu-Fernández, L. Martínez, G. Rodríguez, I. Iglesia, O. Ortiz-Arrabal, et al. Environmental exposure during pregnancy: influence on prenatal development and early life: a comprehensive review. *Fetal diagnosis and therapy*, 48(4):245–257, 2021.

H. J. Hamlin and L. J. Guillette Jr. Embryos as targets of endocrine disrupting contaminants in wildlife. *Birth Defects Research Part C: Embryo Today: Reviews*, 93(1):19–33, 2011.

D. M. Hawkins and C. Weckwerth. Errors in variables regression with value-censored data. *Journal of Chemometrics*, 30(6):332–335, 2016.

C. Hebert, V. Glooschenko, G. D. Haffner, and R. Lazar. Organic contaminants in snapping turtle (chelydra serpentina) populations from southern ontario, canada. *Archives of Environmental contamination and Toxicology*, 24:35–43, 1993.

D. Helsel. *Nondetects and data analysis: statistics for censored environmental data*. Statistics in Practice. John Wiley & Sons, 2005. ISBN 0-471-67173-8.

D. J. Hitchcock, Ø. Varpe, T. Andersen, and K. Borgå. Effects of reproductive strategies on pollutant concentrations in pinnipeds: A meta-analysis. *Oikos*, 126(6):772–781, 2017.

M. S. Humphries, J. G. Myburgh, R. Campbell, A. Buah-Kwofie, and X. Combrink. Organochlorine pesticide bioaccumulation in wild nile crocodile (crocodylus niloticus) fat tissues: environmental influences on changing residue levels and contaminant profiles. *Science of The Total Environment*, 753:142068, 2021.

D. Kim, Z. Chen, L.-F. Zhou, and S.-X. Huang. Air pollutants and early origins of respiratory diseases. *Chronic diseases and translational medicine*, 4(2):75–94, 2018.

G. Kon Kam King, M. L. Delignette-Muller, B. J. Kefford, C. Piscart, and S. Charles. Constructing time-resolved species sensitivity distributions using a hierarchical toxicodynamic model. *Environmental science & technology*, 49(20):12465–12473, 2015.

J. Landes, S. Engelhardt, and F. Pelletier. An introduction to event history analyses for ecologists. *Ecosphere*, 11:e03238, 2020.

A. J. Lawson, C. T. Moore, T. R. Rainwater, F. M. Nilsen, P. M. Wilkinson, R. H. Lowers, L. J. Guillette Jr, K. W. McFadden, and P. G. Jodice. Nonlinear patterns in mercury bioaccumulation in american alligators are a function of predicted age. *Science of the total environment*, 707:135103, 2020.

S. Mikkonen, M. R. Pitkänen, T. Nieminen, A. Lipponen, S. Isokääntä, A. Arola, and K. E. Lehtinen. Technical note: Effects of uncertainties and number of data points on inference from data - a case study on new particle formation. *Atmos. Chem. Phys.*, 19: 12531–12543, 2019.

C. C. Muñoz and P. Vermeiren. Profiles of environmental contaminants in hawksbill turtle egg yolks reflect local to distant pollution sources among nesting beaches in the yucatán peninsula, mexico. *Marine environmental research*, 135:43–54, 2018.

C. C. Muñoz, S. Charles, E. A. McVey, and P. Vermeiren. The attac guiding principles to openly and collaboratively share wildlife ecotoxicology data. *MethodsX*, 10:101987, 2023.

C. Muñoz, S. Charles, and P. Vermeiren. Advancing maternal transfer of organic pollutants across reptiles for conservation and risk assessment purposes. *Environmental Science and Technology*, 58:17567–17579, 2024.

S. Novick, K. Chiswell, and J. Peterson. Bayesian approach to show assay equivalence with replicate measurements over a specified response range. *Statistics in Biopharmaceutical Research*, 4:102–117, 2012.

Pallavi, S. Joshi, D. Singh, M. Kaur, and H.-N. Lee. Comprehensive review of orthogonal regression and its applications in different domains. *Archives of Computational Methods in Engineering*, 29(6):4027–4047, 2022.

X. Qi, S. Zhou, and M. Plummer. On bayesian modelling of censored data in jags. *BMC Bioinformatics*, 23:102, 2022.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL `https://www.R-project.org/`.

A. E. Raftery and S. M. Lewis. [practical markov chain monte carlo]: Comment: One long run with diagnostics: Implementation strategies for markov chain monte carlo. *Statistical Science*, 7(4):493–497, 1992. ISSN 08834237. URL `http://www.jstor.org/stable/2246100`.

P. Reily and H. Patino-Leal. A bayesian study of the error-in-variables model. *Technometrics*, 23:221–231, 1981.

S. Salihovic, E. Lampa, G. Lindström, L. Lind, P. M. Lind, and B. van Bavel. Circulating levels of persistent organic pollutants (pops) among elderly men and women from sweden: results from the prospective investigation of the vasculature in uppsala seniors (pivus). *Environment international*, 44:59–67, 2012.

N. Schuwirth, F. Borgwardt, S. Domisch, M. Friedrichs, M. Kattwinkel, D. Kneis, M. Kuemmerlen, S. Langhans, J. Martínez-López, and P. Vermeiren. How to make ecological models useful for environmental management. *Ecological Modelling*, 411: 108784, 2019.

N. Shoari and J. Dubé. Toward improved analysis of concentration data: embracing nondetects. *Environmental Toxicology and Chemistry*, 37:643–656, 2018.

J. Splett, A. Koepke, and J. F. Estimating the parameters of circles and ellipses using orthogonal distance regression and bayesian errors-in-variables regression. In *JSM Proceedings*, pages 2134–2150, 2019.

J. Tobin. Estimation of relationships for limited dependent variables. *Econometrica*, 26: 24–36, 1958.

P. Vermeiren, C. Lennard, and C. Trave. Habitat, sexual and allometric influences on morphological traits of intertidal crabs. *Estuaries and Coasts*, 44(5):1344–1362, 2021.

L. Wang. Estimation of censored linear errors-in-variables models. *Journal of Econometrics*, 84(2):383–400, 1998.

C. Warwick, P. C. Arena, and C. Steedman. Health implications associated with exposure to farmed and wild sea turtles. *JRSM short reports*, 4(1):1–7, 2013.

S. Watanabe. A widely applicable bayesian information criterion. *The Journal of Machine Learning Research*, 14(1):867–897, 2013.

S. Yu-Sung and Y. Masanao. *R2jags: Using R to Run 'JAGS'*, 2021. URL `https://CRAN.R-project.org/package=R2jags`. R package version 0.7-1.

A. Zuur, J. Hilbe, and E. Ieno. *Chapter2. Generalized linear modelling applied to red squirrel data*, pages 49–88. Highlands Statistics Ltd., Newburgh, United Kingdom, 2015.

7 Supplementary information

S1. JAGS models and data

Model M5 for datasets without censored observations

```
model{

    # Likelihood model
    for(i in 1:n){
        xobs[i] ~ dnorm(x[i], tau)
        yobs[i] ~ dnorm(y[i], tau)
        y[i] <- a + b*x[i]
        x[i] ~ dnorm(mu_x, tau_x)
    }

    # Priors
    tau_x ~ dgamma(1, 0.001)
    tau ~ dgamma(1, 0.001)

    a ~ dnorm(0, 1/pow(sd, 2))
    b ~ dnorm(1, 1/pow(10, 2))
    mu_x ~ dnorm(mean, 1/pow(sd, 2))
    mean <- max / 6
    sd  <- max / 3
}
```

Visualisation of censoring of the dataset

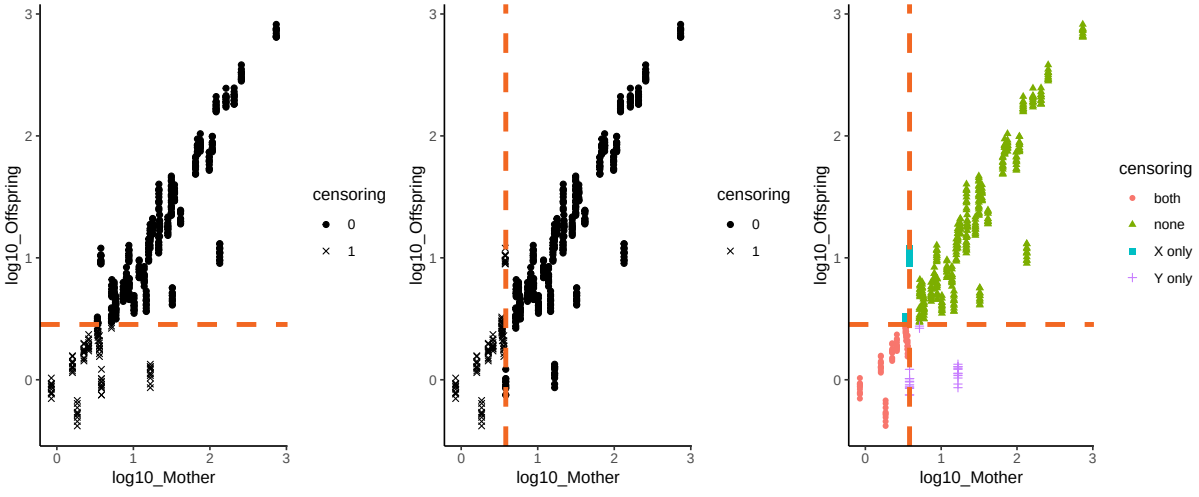


Figure 7: Plots of censored datasets, from left to right: censoring in Y, censoring in X, and censoring in both variables

Model M5 for datasets including censored observations in Y

```
model{

    # Likelihood model - uncensored
    for(i in 1:n){
        yobs[i] ~ dnorm(y[i], tau)
        y[i] <- a + b*x[i]
        x[i] ~ dnorm(mu_x, tau_x)
        xobs[i] ~ dnorm(x[i], tau)
    }

    # Likelihood model - censored
    for(j in 1:c){
        z[j] ~ dbern(p[j])
        p[j] <- pnorm(cut[j], y[n+j], tau)
        y[n+j] <- a + b*x[n+j]
        x[n+j] ~ dnorm(mu_x, tau_x)
        xobs[n+j] ~ dnorm(x[n+j], tau)
    }

    # Priors
    tau_x ~ dgamma(1, 0.001)
    tau ~ dgamma(1, 0.001)

    a ~ dnorm(0, 1/pow(sd, 2))
    b ~ dnorm(1, 1/pow(10, 2))
    mu_x ~ dnorm(mean, 1/pow(sd, 2))
    mean <- max / 6
    sd  <- max / 3
}
```

Model M5 for datasets including censored observations in X and Y

```
model{

    # Likelihood model - uncensored
    for(i in 1:n){
        yobs[i] ~ dnorm(y[i], tau)
        y[i] <- a + b*x[i]
        x[i] ~ dnorm(mu_x, tau_x)
        xobs[i] ~ dnorm(x[i], tau)
    }

    # Likelihood model - censored in Yobs
    for(j in 1:c1){
        z1[j] ~ dbern(p1[j])
        p1[j] <- pnorm(cut1[j], y[n+j], tau)
        y[n+j] <- a + b*x[n+j]
        x[n+j] ~ dnorm(mu_x, tau_x)
        xobs[n+j] ~ dnorm(x[n+j], tau)
    }

   # Likelihood model - censored in Xobs
    for(k in 1:c2){
        yobs[n+c1+k] ~ dnorm(y[n+c1+k], tau)
        y[n+c1+k] <- a + b*x[n+c1+k]
        x[n+c1+k] ~ dnorm(mu_x, tau_x)
        z2[k] ~ dbern(p2[k])
        p2[k] <- pnorm(cut2[k], x[n+c1+k], tau)
    }

    # Likelihood model - censored in Yobs and Xobs
    for(l in 1:c3){
        z3[l] ~ dbern(p3[l])
        p3[l] <- pnorm(cut3[l], y[n+c1+c2+l], tau)
        y[n+c1+c2+l] <- a + b*x[n+c1+c2+l]
        x[n+c1+c2+l] ~ dnorm(mu_x, tau_x)
        z4[l] ~ dbern(p4[l])
        p4[l] <- pnorm(cut4[l], x[n+c1+c2+l], tau)
    }

    # Priors
    tau_x ~ dgamma(1, 0.001)
    tau ~ dgamma(1, 0.001)

    a ~ dnorm(0, 1/pow(sd, 2))
    b ~ dnorm(1, 1/pow(10, 2))
    mu_x ~ dnorm(mean, 1/pow(sd, 2))
    mean <- max / 6
    sd  <- max / 3
}
```

Figure 8: JAGS code for the case of censoring in both variables

S2. Model setup

MCMC setup in Table 2 and 3.

	M6	M5	M4
burn in	60 / 180	5 / 100	2 / 100
Nr. iterations	67428/ 202290	7492 / 22470	7492 / 22470
Thinning	11/10	2 / 10	2 / 10

Table 2: MCMC setup as suggested by Raftery Lewis Diagnostic / effectively used

	M5-censored Y	M5-censored X	M5-censored X and Y
burn in	7 / 100	4 / 100	10/100
Nr. iterations	7492 / 22470	7492 / 22470	14984 / 44940
Thinning	2 / 10	2 /10	4/10

Table 3: MCMC setup as suggested by Raftery Lewis Diagnostic / effectively used

S3. Model convergence

Rhat (Gelman-Rubin diagnostic)

		CV1	CV2	CV3
M6	a	1.019	1.081	1.003
	b	1.02	1.082	1.003
	μ_x	1.001	1.001	1.001
	σ	1.025	1.072	1.004
	σ_x^2	1.01	1.043	1.002
	σ_y^2	1.017	1.069	1.002
M5	a	1.001	1.001	1.003
	b	1.001	1.001	1.002
	μ_x	1.001	1.001	1.001
	σ^2	1.001	1.001	1.002
	σ_x^2	1.001	1.002	1.001
M4	a	1.001	1.001	1.001
	μ_x	1.001	1.001	1.001
	σ^2	1.001	1.001	1.001
	σ_x^2	1.001	1.001	1.001
		Y censoring	X censoring	Y and X censoring
M5 cens	a	1.005	1.004	1.004
	b	1.006	1.003	1.004
	μ_x	1.001	1.001	1.000
	σ	1.001	1.001	1.000
	σ_x^2	1.001	1.000	1.000

Table 4: Rhat values for the different model versions

Effective sample size table

Model M6			Model M5			Model M4			M5 cens		
	mean	sd		mean	sd		mean	sd	Y	X	Y and X
α	93.4	22.1	α	2765.1	56.7	α	6711.0	0.0	379.8	1471.1	497.5
β	89.0	18.3	β	2735.4	57.5				372.4	1604.2	463.9
σ_y^2	698.9	33.9	σ^2	5612.9	727.2	σ^2	6556.6	267.4	3968.8	3472.1	6592.2
σ_u^2	490.9	35.2									
μ_x	27515.0	4150.9	μ_x	6407.8	269.3	μ_x	6910.2	345.0	6711.0	5522.5	12898.9
σ_x^2	238.1	62.5	σ_x^2	6540.2	166.9	σ_x^2	6603.5	186.2	6711.0	6260.2	4422.9

Table 5: Effective sample sizes

S4. Parameter behaviour

Parameter	2.5%	50%	97.5%
α	-5.74	0.00	5.74
β	-18.53	1.02	20.56
τ_y, τ_x, τ_u or τ	25.26	695.24	3684.00
μ_x	-4.31	1.46	7.21

Table 6: Prior parameter distributions

Parameter	Model M6				Model M4			
	2.5%	50%	97.5%	CV	2.5%	50%	97.5%	CV
α	-0.52	-0.39	-0.14	1.03	-0.20	-0.16	-0.13	2.49
β	0.98	1.19	1.22	3.98				
τ_y or τ	9.56	127.55	3224.89	0.01	17.89	22.95	27.84	2.31
τ_u	11.59	16.95	2694.09	0.01				
μ_x	1.14	1.22	1.30	7.96	1.14	1.22	1.30	7.62
τ_x	2.11	2.74	3.47	2.01	1.99	2.39	2.85	2.76

Table 7: Posterior parameter distributions

	Model M5 censored in Y				Model M5 censored in X				Model M5 censored in X & Y			
	2.5%	50%	97.5%	CV	2.5%	50%	97.5%	CV	2.5%	50%	97.5%	CV
α	-0.44	-0.37	-0.30	2.56	-0.36	-0.29	-0.22	1.95	-0.39	-0.32	-0.24	2.11
β	1.11	1.16	1.21	11.43	1.06	1.11	1.16	10.52	1.08	1.13	1.18	10.94
τ	23.80	27.19	30.77	3.94	19.05	21.82	24.61	3.92	21.84	25.18	28.55	3.75
μ_x	1.16	1.22	1.28	10.89	1.16	1.21	1.27	10.89	1.13	1.19	1.25	9.72
τ_x	2.33	2.66	3.03	3.79	2.27	2.62	3.02	3.46	1.95	2.29	2.66	3.19

Table 8: Posterior parameter distributions for censored runs

Posterior correlation matrices

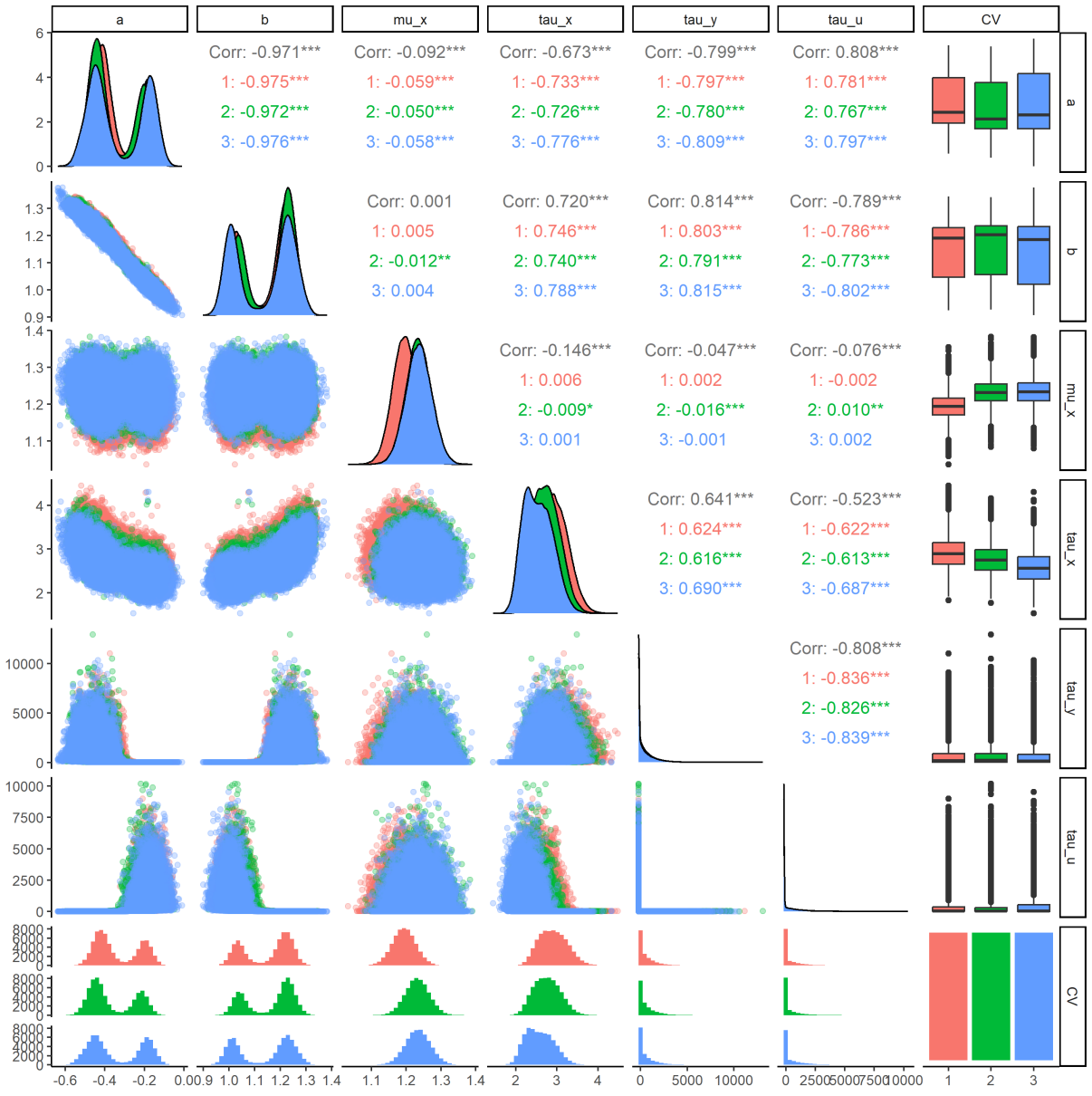


Figure 9: Parameter correlations M6

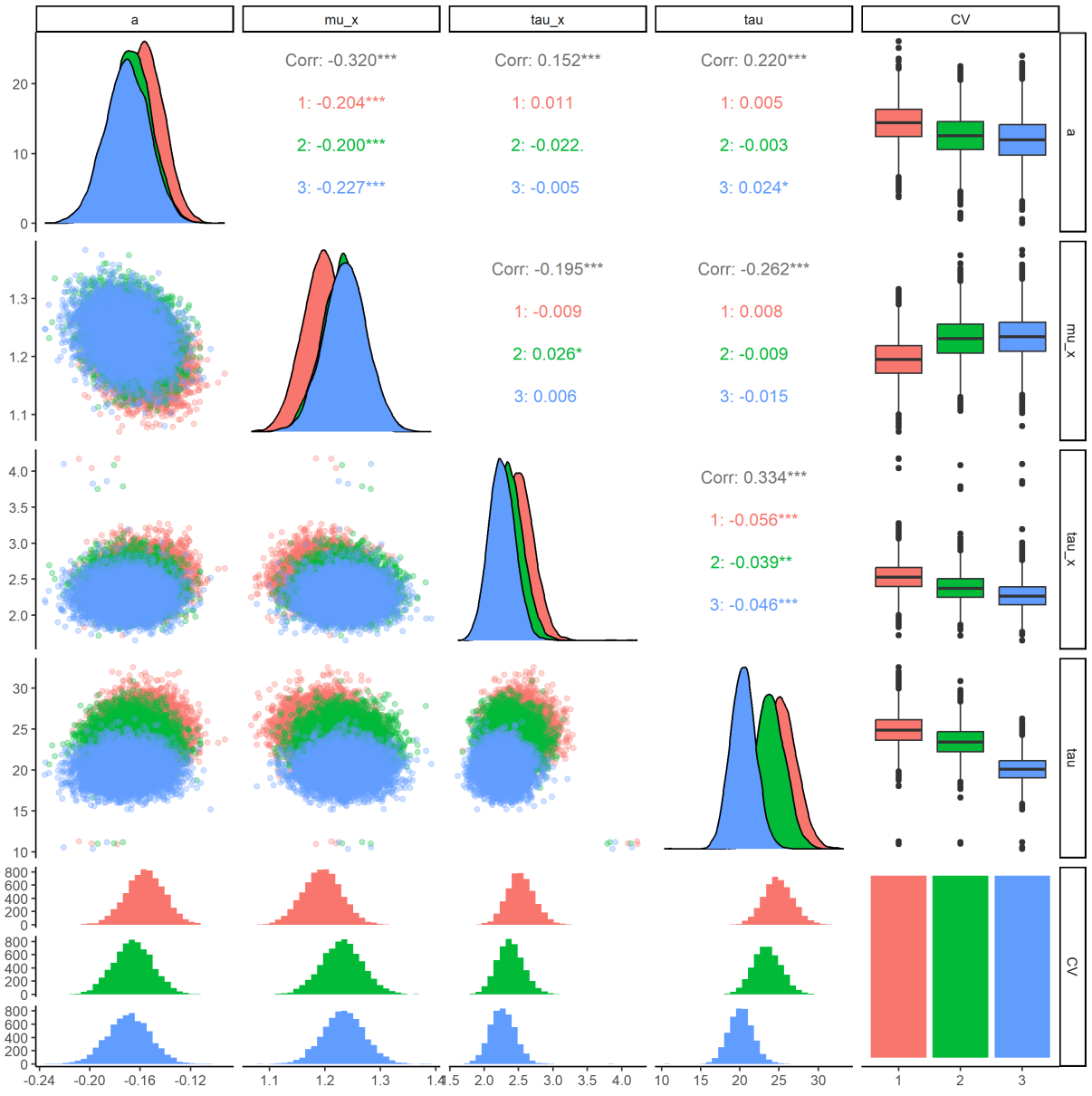


Figure 10: Parameter correlations M4

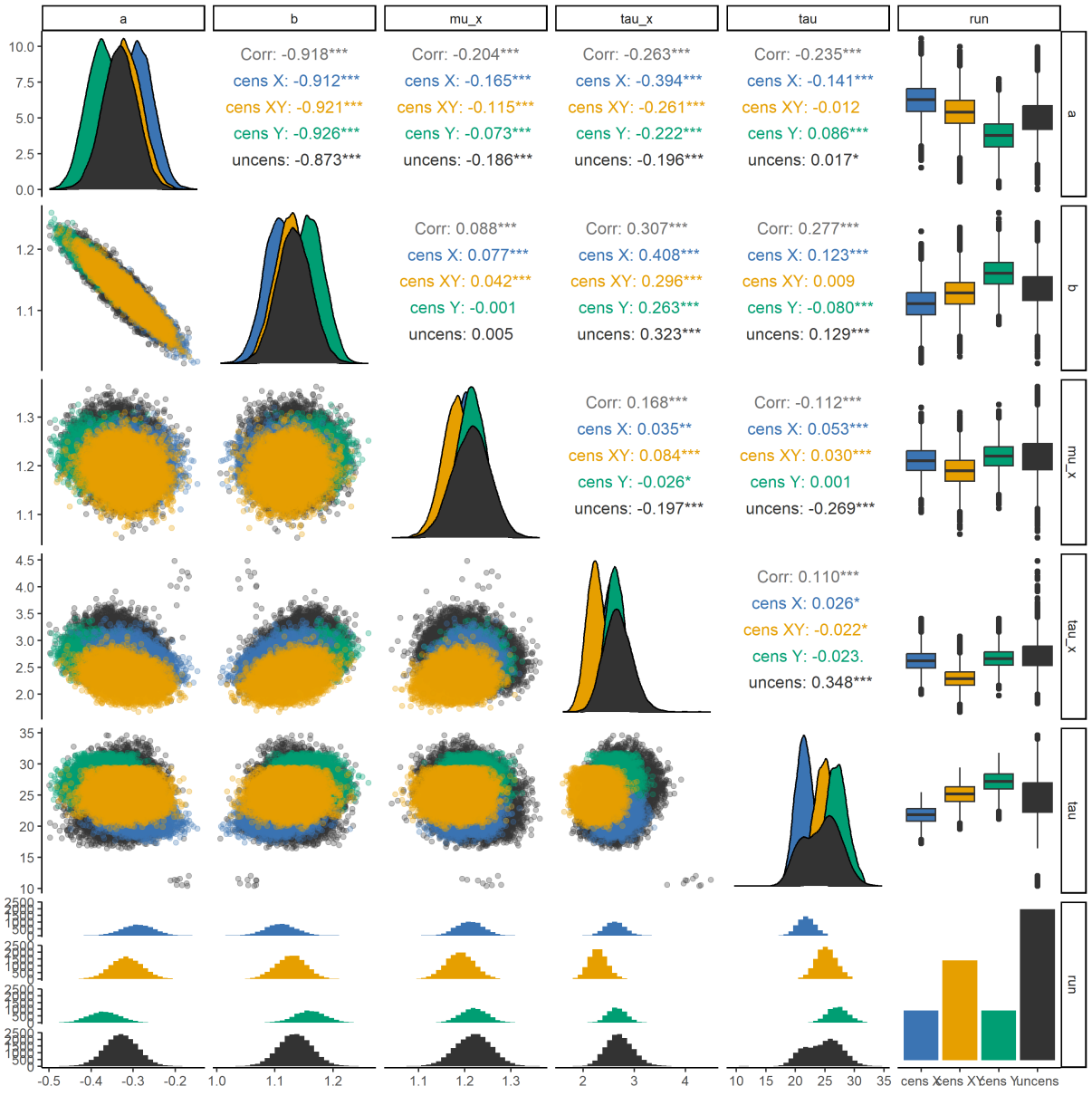


Figure 11: Parameter correlations M5 to censored and uncensored data