

Rebecca Fisher ORCID iD: 0000-0001-5148-6731

David Fox ORCID iD: 0000-0002-3178-7243

Darren Koppel ORCID iD: 0000-0001-7534-237X

Methods for estimating no-effect toxicity concentrations in ecotoxicology.

Methods for estimating no effect toxicity concentrations in ecotoxicology.

Authors: Rebecca Fisher^{1,2*}, David R. Fox^{3,4}, Andrew P. Negri⁵, Joost van Dam⁶, Florita Flores⁵, Darren Koppel¹

Affiliations:

1 Australian Institute of Marine Science, Crawley, WA 6009, Australia

2 University of Western Australia, Crawley, WA 6009, Australia

3 Environmetrics Australia Pty Ltd, Beaumaris, VIC 3193, Australia

4 University of Melbourne, Parkville, VIC 3010, Australia

5 Australian Institute of Marine Science, Townsville, QLD 4810, Australia

6 Australian Institute of Marine Science, Brinkin, NT 0810, Australia

*Corresponding author: Rebecca Fisher

Email: r.fisher@aims.gov.au

Acknowledgement

The authors have no conflicts of interest associated with any of this work. This work was supported through the AIMS research projects 003125 and 003518. We thank Ross Jones, Graeme Batley and Monique Binet for useful comments on the manuscript.

Data accessibility

The data used for analyses in this manuscript are all freely available online. Data for case study 1 can be downloaded at doi:10.4225/15/5746938EC8C8B. Data for case study 2 are available at <https://apps.aims.gov.au/metadata/view/615e208d-f6c7-4b2f-b99e-d4fc50614e48>. In addition, we have made all the data and R code to reproduce the

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ieam.4809.

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

materials in the paper available on github at https://github.com/open-AIMS/NSEC_IEAM.

ABSTRACT

A range of new statistical approaches is being developed and/or adopted in ecotoxicology, that when combined, have the potential to greatly improve the estimation of no-effect toxicity values from Concentration-Response (CR) experimental data. In particular, we compare the existing No-Effect-Concentration (NEC) threshold-based toxicity metric with an alternative No-Significant-Effect-Concentration (NSEC) metric suitable for when CR data do not show evidence of a threshold effect. Using a model averaging approach, these metrics can be combined to yield estimates of no-effect concentrations, and of their uncertainty within a single analysis framework. The outcome is a framework for CR analysis that is robust to uncertainty in the model formulation, and for which resulting no-effect toxicity estimates can be confidently integrated into risk assessment frameworks, such as the Species Sensitivity Distribution (SSD).

Key Points: Estimating concentrations of 'no-effect' of contaminants is critical for effective environmental risk assessment and management. Concentration-Response data can take many forms, and do not always exhibit the clear threshold responses required for No-Effect-Concentration (NEC) threshold-based toxicity models to be validly applied. We compare the existing No-Effect-Concentration (NEC) threshold-based toxicity metric with an alternative No-Significant-Effect-Concentration (NSEC) metric suitable for when CR data do not show evidence of a threshold effect. We show how, embedded within a model averaging approach, the NEC and NSEC can be combined in a single analysis to yield estimates of no-effect concentrations, and of their uncertainty.

KEYWORDS

Concentration-Response modelling; toxicity estimate; No Effect Concentration; Ecosystem protection; statistical ecotoxicology

INTRODUCTION

The primary concern under most environmental regulatory risk assessments is whether there is a level of toxicity that poses a risk to the survival of a population.

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

Generally, within a protective risk management framework the goal is to determine the concentration at which there can be considered no adverse effects on the ecosystem. For this reason, estimation of no-effect concentrations for substances of concern are integral to both risk assessment regulation and for the establishment of environmental quality standards (de Bruijn and Hof 1997). In several jurisdictions the species sensitivity distribution (SSD, Posthuma et al. (2001)) is used to extrapolate, from individual estimates of safe concentrations for a range of species, the percentage of the community that is protected at concentration “x” (PC_x), or hazardous concentrations (HC_x) for 100-x% of all species, that are applied in most current formal water quality guideline value (GV) derivations (Fox et al. 2021). The data underpinning the SSD are the toxicity estimates extracted from individual species concentration-response (CR) experimental data. Ideally, if the goal is high confidence in the level of protection achieved, the input toxicity data should represent the maximum concentration of no “effect” for each species.

There is a wide range of statistical metrics that have been adopted to estimate no-effect toxicity values for use in SSDs (Table 1), and for pragmatic reasons a range of metrics is generally allowed within most regulatory frameworks. These include the No-Observed-Effect-Concentration (*NOEC*); Effect-Concentration of a defined percentage *x* (typically 10%, *EC10*, *LC10*) and the No-Effect-Concentration (*NEC*, Table 1). There is considerable debate in the literature regarding the validity and value of different approaches (Fox 2008; Warne and Van Dam 2008; Fox 2012; Green et al. 2013; Fox and Landis 2016a, 2016b). All three methods are used in practice with each having strengths and weaknesses across the myriad of situations that arise in CR modelling. The current Australian guidelines state that: ‘... *the preferred order of statistical estimates of chronic*

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

toxicity to calculate default and site-specific GVs is: chronic NEC, EC/IC/LCx where $x \leq 10$, BEC10, EC/IC/LC15–20, and NOEC. While all of these acceptable statistical estimates of toxicity are not numerically the same, they are all treated as equivalent for the purposes of deriving GVs ...’. However, for SSDs developed for the purpose of estimating protective guideline concentrations, the recommendation that “effect” concentrations are allowable in an SSD could be considered problematic because, by definition, such toxicity measures represent an “effect” of the specified amount, and the resulting SSD cannot be assumed to be protective (Fisher and Fox 2023). It is for this reason that an NEC is the preferred toxicity metric for SSD modelling in Australia (Warne et al. 2018).

The NEC is typically estimated using a threshold model (Pires et al. 2002; Fox 2008; Fox 2010), and represents the maximum concentration for which there is no response for a given species, thereby providing a toxicity measure that is ideal for incorporation into SSDs aimed at estimating protective concentrations of contaminants. While the NEC is considered the preferred measure for inclusion into SSDs in at least some jurisdictions (Warne et al. 2018), CR data do not necessarily exhibit abrupt threshold-like responses and applying a threshold model in this case will lead to poor outcomes (Krull 2020). This phenomenon has occurred in our own work, where at times attempts to fit a threshold model has failed using standard packages such as `drc` in R (Ritz et al. 2015), or when it does fit successfully, may yield values that are higher than even the EC10 (Negri et al. 2021). In this case, a threshold model of response is clearly not appropriate, and the NEC will not be a suitable estimate of the no-effect concentration.

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

More recently Fisher and Fox (2023) introduced the No-Significant-Effect Concentration (NSEC) as an alternative to the NEC when a threshold is either untenable or not indicated by the data. Initially proposed by (Bellio et al. 2000) and further developed by (Chèvre et al. 2002). Full details of the method are provided in Fisher and Fox (2023) and are not repeated here. Briefly, the technique involves fitting a non-linear model to the CR data and evaluating the concentration above which the difference between the predicted value of the response at the control and that concentration is statistically significant, based on the estimated variability and a pre-defined level of statistical significance. The NSEC represents an improvement over the conceptually linked no-observed-effect-concentration (NOEC) often used in ecotoxicology by decoupling the estimate from the treatment concentrations specifically used in a given experiment and allowing an estimate of precision (Fisher and Fox 2023). While Fisher and Fox (2023) provide a thorough description of the mathematical derivation of the frequentist version of the NSEC, as well as implementation of the Bayesian equivalent, the concept has yet to be evaluated more extensively using simulated and case study data.

In addition to the range of toxicity estimates that are used in SSD modelling, there is an even wider range of possible CR models that may be used to derive them. For example, the popular frequentist CR modelling R package *drc* (Ritz et al. 2015) contains 23 non-linear functions that can be used, as does our recently developed Bayesian CR modelling package *bayesnec* (Fisher et al. 2021). These include multiple threshold models, as well as a wide array of models representing a smooth decline with increasing toxicant concentration. While there may be underlying physiological and toxicological mechanisms suggesting a threshold model is appropriate in some cases (Chapman and

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

Wang 2000), this is not always known and clear thresholds may not always be apparent in the observed response relationship due to complex physiological responses and variable individual tolerances resulting a wide tolerance distribution across a population (Van Straalen 1997). Without a theoretical basis for model selection, it becomes difficult to a-priori determine the best model to use for a given set of data. Given the many possible curves that may be used to describe the data, model uncertainty can be very high and model selection ambiguous when many of these appear to fit the data equally well. Model averaging (Burnham and Anderson 2002) represents an approach that can provide a robust way of accommodating model uncertainty. Model averaging involves fitting a candidate set of plausible models to the data and obtaining weighted averaged estimates of the metrics of interest (Burnham and Anderson 2002). Weights are based on the relative fit of each model to the data. Model averaging is currently being considered as a potentially more robust framework for SSD modelling (Fox et al. 2021) and is widely used in ecology (Dormann et al. 2018).

Here we explore estimation of no-effect toxicity values using the recently introduced NSEC toxicity measure (Fisher and Fox 2023) in combination with NEC, within a model averaging framework. We begin with a simplified simulation study including only two alternative models, an NEC threshold model and a simple sigmoidal decay function. We then present a case study using real data showing how the NSEC and NEC values, when combined using model averaging can yield estimates of no-effect concentrations, along with estimation of their uncertainty within a single analysis framework.

SIMULATION STUDY

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

We constructed a simulation study to compare toxicity estimates obtained using NEC, EC_x and the NSEC, across a range of experimental designs with differing replication. Simulations were based on four alternative scenarios representing four different theoretical concentration response relationships of a binary response endpoint. For all scenarios, we assume a data structure of the following form: $\{y_i = \text{number of 'successful' outcomes out of } n_i \text{ replications of an experiment at concentration } x_i\}$. Our response y_i is modelled using a binomial distribution, with expected value μ_i , given the number of binomial ‘trials’ n_i :

$$y_i \sim \text{bin}(\mu_i, n_i)$$

where for the first two scenarios the expected value μ_i was predicted using the three-parameter NEC exponential decay threshold model of Fox (2010), with the parameters α (top, y intercept), β (exponential decay rate) and γ (the NEC threshold, see equation 1).

$$\mu(x_i; \alpha, \beta, \gamma) = \alpha \exp[-\beta(x_i - \gamma)I(x_i; \gamma)] ; \{\alpha, \beta, \gamma\} > 0 \text{ and}$$

the indicator function $I(x_i; \gamma)$,:

$$I(x; \gamma) = \begin{cases} 1 & \text{if } x > \gamma \\ 0 & \text{otherwise} \end{cases} \quad I(x; \gamma) = \begin{cases} 1 & \text{if } x > \gamma \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The second two scenarios were based on a three-parameter sigmoidal decay model representing a smooth decline with concentration (see equation 2). This includes the same parameters α (top, y intercept), β (exponential decay rate), and an additional parameter δ influencing shape of the decay function.

$$\mu(x_i; \alpha, \beta, \delta) = \alpha \exp[-\beta x_i^\delta] \quad \{\alpha, \beta, \delta\} > 0 \quad (2)$$

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

These models were used to generate predicted data for a given concentration (x), using the two R functions given in the Supplementary Information.

Parameters were selected for α , β , γ and δ (Table 2) to yield the four curves shown in

Figure 1A. Simulated data sets were randomly generated using these four curves based on a binomial distribution, to represent a hypothetical binary endpoint (e.g. survival) with a mean value of 90% success for the control (α - top,). This was achieved by generating theoretical predicted probabilities for two theoretical experimental treatments sequences with either 8 (low density design) or 12 (high density design) treatment concentration values distributed evenly from 0.01 to a maximum hypothetical value of 10 concentration units. We applied the base R function *rbinom()* to randomly simulate binomial response data at each treatment level for the predicted probability, with varying levels of n (5 or 10 *replicates*) and size (10 or 20 binomial *trials*, representing the number of individual test organisms within each replicate). The number of trials used was based loosely around our experience with some of the commercial accredited ecotoxicological tests offered in Australia, such as the *S. echinata* (100 individuals/replicate, <https://www.ecotox.com.au/wp-content/uploads/2018/11/Testfactsheet7.pdf>) *E. mathaei*, WIECX-25-Sea Urchin Larval, Test ID: 69669, 200 individuals/replicate) larval development tests. The combination of experimental conditions resulted in 8 different sets of simulated data, for each of the Scenarios examined, and representative examples are shown for each scenario are in Figure 1B. This simulation process was repeated 100 times for each design, for all four model scenarios.

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

We used the R package jagsNEC <https://github.com/open-AIMS/NEC-estimation> (Fisher et al. 2020) to fit both the 3 parameter NEC and sigmoidal models to the randomly simulated data. Both models were fit in a single call to `fit.jagsMANEC`, with `model.set` set to `c("NEC3param", "ECxsigmoidal")`, which represent the two original model types used to generate the two NEC and sigmoidal scenarios (see equation 1 and equation 2). Posterior estimates for the NEC were obtained directly from the NEC model fit for each simulated dataset as the estimated η parameter. Posterior estimates for NSEC were obtained with the function `extract_NSEC` (Fisher et al. 2020) which implements the Bayesian method for estimating No-Significant-Effect-Concentrations, as described in Fisher and Fox (2023), and for the simulation study was based solely on the estimates obtained from the smooth sigmoidal curve. In addition to the NEC and NSEC estimated from their respective threshold and sigmoidal models, a model averaging approach is used to provide a combined estimate. This model averaging strategy uses an information theoretic approach (Burnham and Anderson 2002), and weights the average of the NEC and NSEC based on the relative support of each model given the data, providing a combined estimate of no-effect, which we denote as the N(S)EC. Within this framework, if the NEC model(s) fits the data better these will have high weight and the resulting estimate' will be a true NEC estimate. Conversely, if the sigmoidal model(s) fit better, the resulting toxicity estimate will be based largely on a NSEC estimate. We also similarly obtained model averaged posterior estimates of EC10, EC5 and EC1 using the function `extract_ECx` (Fisher et al., 2020).

This article is protected by copyright. All rights reserved.

Accepted Article

Revised manuscript 10th July 2023

Comparing model weights

The jagsNEC package uses deviance information criterion (DIC, Spiegelhalter et al. (2002)) based model weights to generate model averaged estimates of the toxicity estimates (e.g. EC10 and NSEC). To compare N(S)EC estimates across the different scenarios, we first wanted to establish if the model weighting procedure works effectively in these examples.

We found that the underlying generating model usually had the highest weight for most simulated datasets (

Figure 2). There were some exceptions, with a few simulations resulting in high weight for the model not used to generate the data, particularly for low sampling and treatment density (

Figure 2A).

There was a tendency for the data generated from the second sigmoidal scenario to yield a higher weight for the NEC model, particularly when there are few treatments (Sigmoidal 2,

Figure 2). This tendency persists, even with quite high replication within treatments. There was also a tendency for data generated using the first NEC scenario to have a high weight for the sigmoidal model when sampling density is lower, with this occurring more frequently when there were few treatments (NEC 1,

Figure 2).

Increasing the number of treatments had a more profound effect on improving the relative weights of the two competing models, than increasing the number of trials or replicates within each treatment (Figure 2B). For example, the weight for the NEC model for the NEC 1 scenario was 0.987 for the design with the fewest replicates (5 replicates each with 10 trials) and 12 treatments (a total of 600 trials, Figure 2B, Table S1), which is similar to a design with 8 treatments and double the level of within treatment replication (5 replicates each with 20 trials, a total of 800 trials, weight = 0.984, Figure 2, Table S1).

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

For the Sigmoidal 2 scenario, this effect was even more pronounced, with median weights better for the design with 12 treatments across all levels of replication compared to 8 treatments (Figure 2).

Comparing toxicity estimates

We examined the NSEC, NEC and N(S)EC values estimated using jagsNEC (Fisher et al. 2020) for all the simulation scenarios examined, along with EC1, EC5 and EC10, to see how these compared to each other (

Figure 3, Table S2). The model averaged estimated N(S)EC values are very close to the true NEC estimates for data simulated using an NEC-type model, although the NSEC estimates are considerably lower than the NEC (NEC 1 and NEC 2,

Figure 3A). The close agreement between NEC and N(S)EC reflects the high weight that the NEC model has for the simulated data generated from an underlying NEC model, and the NEC model's greater overall contribution to the combined N(S)EC estimate.

Model averaged estimated N(S)EC values are very similar to the NSEC estimates for data simulated using a sigmoidal model, and for both sigmoidal scenarios the NSEC estimates are considerably lower than an NEC estimated from these data (Sigmoidal 1 and Sigmoidal 2,

Figure 3A). For sigmoidal models there is no true theoretical NEC and we can only compare to true EC_x estimates. For the sigmoidal scenarios, the estimated N(S)EC and NSEC values fall across the range of EC1 to EC10 (Sigmoidal 1 and Sigmoidal 2,

Figure 3A).

When NEC is estimated by fitting the NEC threshold model to data generated using the two sigmoidal models (Sigmoidal 1 and Sigmoidal 2), estimated values are quite high relative to even the EC10 estimate, with most estimates higher than the true EC10 value (

Figure 3). The estimated N(S)EC and NSEC values are always lower than the estimated NEC value for these sigmoidal models (

Figure 3).

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

Estimates of all toxicity values (including NEC, NSEC and ECx) are more variable with lower sampling effort across all four scenarios (Figure 3B). Estimates of NSEC as well as ECx show greater variability with low sampling effort, and for some scenarios variability is further reduced for a design with 12 compared to 8 treatments for a similar sampling effort (blue line is often lower than the green line, Figure 3B).

Actual effects of estimated toxicity

For our simulation study, the underlying response-generating which allows estimation of the actual effect size for the model averaged N(S)EC, NSEC and NEC estimates, as well as the estimated EC1, EC5 and EC10 values (Figure 4, Table S2).

For datasets generated using the two NEC models, estimates of NEC, NSEC and the combined N(S)EC were all very close to the theoretical expectation of 0% effect, with a mean difference from expected usually at zero, even for designs with relatively low sampling effort (Figure 4B). The exception is the NEC estimate for the second NEC scenario for the design with the least sampling effort, although even then the mean estimated effect is still only ~3% (Figure 4B). This reflects the relative robustness of the NEC as a measure of no-effect toxicity for data that have a distinct threshold effect and this robustness appears to be retained when the NEC is combined with the NSEC in a model averaging framework. The ECx estimates derived from models fitted to datasets generated using the two NEC-threshold models tended to be lower than expected, particularly for designs with low sampling effort (Figure 4B). There was a somewhat complex pattern for both EC05 and EC10 estimates with the sampling design for the second NEC scenario, possibly due to the position of the treatment concentrations relative to the theoretical position of the NEC (Figure 4B).

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

For data simulated using the two sigmoidal models actual effect sizes of all toxicity estimates varied substantially across simulations, and all were influenced by sampling density (Figure 4). The NEC values estimated from data generated using sigmoidal models represented relatively high actual effect sizes (15 to 40% depending on the design) and did not necessarily improve with greater sampling effort (Figure 4B).

There was substantial variability in the actual effect size of the estimated NSEC and N(S)EC values (Figure 4A). Estimates for the NSEC and N(S)EC represented effects nearing 10% on average for designs with the poorest sampling effort, but the calculated effects decreased substantially with increasing sampling effort, representing an effect of ~1% on average for the most replicated design in our simulations (Figure 4B). Like the NSEC, there was also substantial variability in the actual effect size of the estimated ECx values, with ECx estimates tending to be greater than the theoretical value when sampling effort is low, particularly for data based on the second sigmoidal scenario (Figure 4B). As sampling effort increases, this bias gradually reduces, and mean estimates correspond to their true value for the designs with the greatest sampling effort (Figure 4B). For the second sigmoidal scenario in particular, increasing the number of treatment concentration reduces bias in estimates more quickly for a given sampling effort (Figure 4B).

CASE STUDIES

Case study 1

In our first case study we demonstrate how the “no effect” copper toxicity to the Antarctic marine microalgae *Cryothecomonas armigera* can be simply estimated using model averaging across a range of NEC threshold and smooth sigmoidal curves. The data

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

for this case study are available to download from doi:10.4225/15/5746938EC8C8B and describe the decreasing microalgal population growth rate (normalized as a percentage of controls) following increasing exposure to dissolved copper (Koppel et al. 2017). As the growth rate data were collated across several separate concentration experiments, to ensure alignment they were normalized to the maximum growth observed within each experiment. Graphical inspection of the data suggested that there was a natural upper bound to the growth rate. To accommodate a response variable with fixed lower (0 growth) and upper (maximum growth) bounds we used a re-parameterization of the Beta distribution (equation 3) in the likelihood function, following normalization relative to the maximum. The likelihood function for the mean response (μ) is estimated as a function of the two shape parameters of the Beta distribution, using the latent parameter, phi (ϕ):

$$y_i \sim \text{Beta}(s_{\text{ape}1_i}, s_{\text{ape}2_i})$$

$$s_{\text{ape}1_i} = \mu_i \phi$$

$$s_{\text{ape}2_i} = (1 - \mu_i \phi) \quad (3)$$

Models were fit using the same jagsNEC package (Fisher et al. 2020) as used in the simulation study. This package uses a model averaging approach based on deviance information criterion (DIC) weighted averaged predictions across a potential candidate model set composed of a range of functional NEC models adapted from (Fox 2010), including two NEC models (NEC3param, NEC4param) and a range of commonly used sigmoidal models (ECx models: ECxLinear, ECx4param, ECxExp, ECxSigmoidal, ECxWeibull1 and ECxWeibull2, see Fisher et al. (2020) and the supplementary

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

information for more details on the included models, including model formula). We used the default settings in the function `fit.jagsMANEC`, with 5000 iterations as burn in, 10000 update iterations and three chains. In all cases, chain mixing was assessed visually, and where models showed poor mixing, they were excluded from model averaged estimates of toxicity estimates.

The DIC based model weights for the resulting CR curves were spread relatively evenly across three smooth models, ECx4param, ECxsigmoidal, and ECxWeibull2 (Table 3, Figure 5). The NSEC estimates for these models were all comparable at 6 (1-11), 7 (3-11), and 6 (1-11) $\mu\text{g L}^{-1}$ (Figure 5). We denote the model averaged “no effect” concentration as the N(S)EC, to indicate it is a weighted average of both NEC and NSEC values. The NEC models had an averaged NEC of 4.6 (3.9-4.9) $\mu\text{g L}^{-1}$ (Table 4) but low weights (0-0.001, Table 3), so the resulting N(S)EC of 7 (1-11) $\mu\text{g L}^{-1}$ reflects the weight of the smooth models (Figure 5). While the N(S)EC value is lower than the NEC, NOEC or the EC10 (Table 4), the values are slightly higher than the EC1 estimate (Table 4). The estimated “effect” of the N(S)EC value is ~2.3 % (Table 4).

Case study 2

In our second case study, we use data from our recent publication (Negri et al. 2021) that was also originally analysed via the `jagsNEC` package (Fisher et al. 2020). These data examined the responses of eight tropical marine species to the water accommodated fraction of gas condensate (light crude oil) from the Ichthys and Prelude gas fields off the tropical northwest coast of Australia. The main aim of the original study was to build an SSD following the standard guideline methods (Warne et al. 2015; Warne et al. 2018) to validate mixture toxicity modelling for petroleum hydrocarbons. In their

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

original analysis either the EC10 or the NEC was used for input into the SSD (whichever was more conservative), as both of these are considered valid toxicity estimates for the purpose of developing SSDs (Warne et al. 2018). Here we revisit the analysis of the underlying CR curves and use these to explore issues with fitting NEC models to real data, as well as how the N(S)EC toxicity estimate performs in practice. The original data considered eight species. However, three of these did not show a complete response at the highest concentration examined (*Acropora muricata*, *Phyllospongia foliascens*, and *Rhodomonas salina*), and fourth exhibited evidence of hormesis (*Cladocopium goreauii*) (Negri et al. 2021). For simplicity, these four species were excluded here, leaving the remaining four species - *Acropora millepora* (coral), *Stomopneustes variolaris* (sea urchin), *Nassarius dorsatus* (gastropod) and *Amphibalanus amphitrite* (barnacle) to be used in our case study assessment.

Substantial details regarding the collection of these data including the experimental conditions and a description of each assay are described in the original study (Negri et al. 2021) and the supplementary files <https://ars.els-cdn.com/content/image/1-s2.0-S0025326X21009334-mmc1.pdf>. Response data for assays based on a decline in the percentage of larvae successfully completing metamorphosis (*A. millepora* and *A. amphitrite*) or fertilization success (*S. variolaris*) were initially fit using a binomial likelihood function. However, in all cases, these initial models were over dispersed. Instead, these were converted to a proportion and the data fit using a Beta likelihood function, which allows for a more flexible relationship between the mean and the variance relative to the binomial distribution. Response data for assays based on specific growth rate were normalised if values exceeded one (*N. dorsatus*) by

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

dividing by the maximum observed value (to reflect proportional decline) and then also modelled using a Beta distribution, as there were similarly upper bounded as overserved for case study 1. Concentration data were included in all models on a log scale, as this was the natural scaling evident in the placement of treatments across the concentration range considered. All models were fit using the R package jagsNEC (Fisher et al. 2020). In all cases, chain mixing was assessed visually, and where models showed poor mixing, they were excluded from model averaged estimates of toxicity estimates. As for case study 1, only models with NEC as a specific parameter were used to obtain a DIC weighted model averaged estimate of NEC (i.e., both models with the prefix NEC). Estimates of 1% and 10% effect (EC10) and N(S)EC were calculated using DIC weighted model averaged estimates obtained from all successfully fitted models (both models with a prefix NEC and models with the prefix ECx, see Fisher et al. (2020)), although in the latter case N(S)EC is a weighted averaged of the NEC and NSEC estimates of the underlying models.

The highest weighting model varied across the four species examined (Figure 6). For *A. amphitrite* there was strong weight for a single model – which was the original NEC 3-parameter model (Fox 2010) (Figure 6). For this species the model average based on “all” models was essentially identical to the NEC model set, and the estimated values for the NEC, EC10 and N(S)EC were all very similar (with NEC marginally lower than the other two estimates), although the confidence bands for the NEC were substantially narrower than for the other two toxicity estimates (Figure 6, Table 4). For the other three species there was substantial support for more than one model, and this generally included support for one of the “NEC” model types, as well as a smooth “ECx” model

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

type (see Figure 6). For *N. dorsatus*, the EC10 estimate is nearly identical to the estimated NEC (blue vertical line overlaps the red vertical line), with the estimated N(S)EC only slightly less, with overlapping confidence intervals (Figure 6, Table 4). *A. millepora* also had very similar estimates for the N(S)EC, EC10 and NEC, which all have overlapping confidence intervals, although the N(S)EC was the most conservative and NEC the least conservative for this species (Figure 6, Table 4). For *S. variolaris* the NEC estimate was definitively higher than the estimated EC10, and the N(S)EC estimate was lower than either the NEC or EC10 (Figure 6, Table 4). As the N(S)EC estimate was quite low for *S. variolaris*, in order provide context we calculated the NOEC value also using a Beta distribution, fitting the concentration data as a treatment factor via the package glmmTMB (Brooks et al. 2017) and doing a Dunnett's test via glht from the package multcomp (Hothorn et al. 2008). We found that the NOEC value was 31.8 $\mu\text{g L}^{-1}$, which is far more conservative compared to our estimated N(S)EC (140 $\mu\text{g L}^{-1}$, Table 4), providing further support that an EC10 estimate for these data would be highly non-conservative, representing a potentially environmentally significant effect. This also shows that while the estimated N(S)EC is quite low, and may appear to some overly conservative, it is less conservative than a calculated NOEC based on the same data. The estimated “effect” of the N(S)EC value for this species is ~1.5 % (Table 4).

DISCUSSION

The estimation of No-Effect concentrations is critical for deriving safe concentrations of thresholds in the environment. Threshold-based models for estimating toxicity estimates of no effect are ideal because they directly estimate the no effect concentration as a parameter in the model. However, some concentration-response data

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

show no threshold effect but instead exhibit a gradual decline with concentration rendering an NEC model inappropriate (Fisher and Fox 2023). Indeed, for both our simulation and the case study examples, when the 3-parameter NEC model of Fox (2010) is fit to smooth sigmoidal data the resulting estimates of NEC are often higher than even the EC10, with simulations suggesting true effects as high as 30%. This occurs because the 3-parameter NEC model fit is unable to capture the gradual decline inherent in smoothly sigmoidal data and reinforces the case that fitting threshold models to smoothly sigmoidal data is clearly not appropriate (Fisher and Fox 2023). In the original analysis of the second case study data, the authors adopted the conservative approach of selecting the lower value of either the NEC or EC10 (estimated using models properly capturing the smooth decline) for inclusion in the final SSD (Negri et al. 2021). This decision is consistent with the current Australian guidelines that aim to derive concentrations that are unambiguously protective of the whole community (Warne et al. 2018). Certainly, using the NEC in this case would result in the derivation of thresholds that are unlikely to be as protective of the community as intended.

The fact that NEC models provide a poor fit to smoothly sigmoidal data resulting in NEC estimates higher in some cases than the EC10 is to be expected, because the NEC three-parameter model fitted here was not used to generate the underlying sigmoidal simulation data. However, the general lack of alternative NEC step models means that this is exactly how the NEC would be estimated from these data in practice. We attempted to resolve this issue when it arose in the analysis of case study 2 data by including an additional model that expanded the original three-parameter model of (Fox 2010) to allow a sigmoidal decline in the response in the development of jagsNEC

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

(Fisher et al. 2020). However, applying this model to real data often resulted in highly unresolved NEC estimates with extremely wide confidence bands (see Supplementary Information, Figure S1). There may be value in further developing the NEC modelling framework to allow for sigmoidal declines using other threshold model parameterisations. However, it seems likely that similar unresolved NEC estimations will result from any sigmoidal model with a relatively flat upper asymptote as the transition point (step/threshold).

For simulated data based on the underlying NEC model the model averaged estimated N(S)EC values are very close to the true NEC estimates for data simulated using an NEC-type model, although the NSEC estimates are considerably lower than the NEC. The high level of similarity between NEC and N(S)EC reflects the high relative weight that the NEC model has when fit to data generated from an underlying NEC-threshold model, which results in the NEC model's greater overall contribution to the combined N(S)EC estimate. Model averaging alleviates the need to undertake separate analyses (threshold versus non-threshold models), providing a method for estimating a no-effect concentration within a single analysis framework.

Our simulation study showed that for well-designed experiments with a large number of treatment concentrations the DIC model weights we used here were usually highest for the underlying data generating model. This suggests the DIC weights did a reasonable job of resolving the true model, and that the model average inference is quite robust, at least for these simulations study data. Weights did vary more widely for simulations with fewer treatment concentrations (8 rather than 12), re-enforcing the idea that replication within treatments (a requirement of the one-way ANOVA methods used

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

to generate NOECs) should be reduced in favour of increasing the number of treatments (Fox 2016).

The DIC weights used here are simple to extract from jags models which provide DIC by default. These are generally considered analogous to other information theoretic metrics such as AIC (Spiegelhalter et al. 2002). However, there are potentially much better weighting strategies available for Bayesian model fits (Gelman et al. 2014). This was one motivation for the development of the bayesneec package (Fisher et al. 2021) based on brms (Bürkner 2017), for which the loo (Vehtari et al. 2020) package provides a range of weighting options (Vehtari et al. 2017).

Model averaging provides additional benefits to ecotoxicologists because it avoids having to decide which model to use when there is no overarching theoretical basis for model selection. Perhaps more importantly, model averaging eliminates the need to select only a single model for inference. This is helpful when more than one model fits the data well, such as in Case Study 1 where model weights were broadly equivalent for three competing models, although in this case the estimated curves (and resulting inference) are very similar (

Figure 4). We have come across more problematic examples when the log-logistic and Weibull type 1 and 2 models from drc (Ritz et al. 2015) can result in different effects estimates, particularly at the EC10 level, despite having similar AIC values and associated weights. Model averaging ensures that estimated toxicity values are robust, defensible, and usable within the relevant decision context or regulatory framework. Note that we are not suggesting here that model averaging should encourage uncritical “automated” analysis – indeed careful selection of the candidate model set, the appropriate application of statistical methods (see below), and critical evaluation of the model outputs will remain essential elements of the analysis of a CR experiment.

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

Within the model averaging framework being put forward here, we suggest that the NSEC (as calculated from sigmoidal models, Fisher and Fox 2023) be combined with the NEC estimated from threshold models to estimate an overall no-effect concentration - the N(S)EC. We believe that the NSEC is a more appropriate estimate of “no-effect” than a low- effect EC_x estimate to use within this model averaging framework, because an EC_x – by definition, represents some level of effect. Because the NSEC estimate is based on a significance test relative to the controls (Fisher and Fox 2023), these estimates will have some level of non-zero observed “effect” associated with them, in the same way that the No Observed Effect Concentration (NOEC) can also represent some level of effect (see Mebane et al. (2008)). Van Der Hoeven (1997) argues that because there is a minimum non-zero effect size corresponding to any given NOEC, setting defined effects “x” (of similar size to that minimum) should therefore be acceptable. However, the difference is that the “x” in EC_x is typically arbitrarily defined, with no direct link to the biological implications of that effect. The “x” in a NOEC (or NSEC) is defined as zero, with any deviation from zero being regarded as random noise. While it is true that the probability level considered in the estimation of NOEC is often arbitrary (and perhaps should be more carefully considered with respect to the specific decision context (see for example the probabilistic thresholds in Fisher et al. (2018))), the basic mechanism of hypothesis testing at the heart of the NOEC (and NSEC) at least provides a clear and coherent framework with which to make decisions under uncertainty. Indeed both Van Der Hoeven (1997) and de Bruijn and Hof (1997) point out the fact that selection of an appropriate “x” is problematic from a policy perspective.

This article is protected by copyright. All rights reserved.

Accepted Article

Revised manuscript 10th July 2023

Where there is sound ecological or biological justification with which to set a meaningful “essentially no-effect” value (x) for an EC_x , the use of EC_x may be justified in the derivation of protective guidelines. However, there are very few situations in which it is possible to theoretically define such an “ x ”, and “ x ” is typically set at an arbitrary, small, fixed value. A fixed low level of effect as defined by an EC_x may not translate into a meaningful biological effect. Alternatively, the fixed value of “ x ” may represent a very large effect. Green et al. (2013) discusses a range of examples where statistically significant effects can occur well below a 10% effect, and such effects are likely of biological importance. In this case, using an EC_{10} may result in highly non-conservative toxicity estimates that will not provide an adequate surrogate for a true NEC. While the NSEC can have an “effect” in the sense that it represents a response value that is less than the mean level of the response for the control, this effect is defined as non-significant in the context of the variability observed under the experimental conditions. If the “effect” allowed remains within the bounds of the naturally occurring values in the absence of the toxicant, it seems reasonable to infer such effects are unlikely to have little ecological/biological consequence on the population.

It could be argued that the application of NEC and NSEC over EC_x values in constructing SSDs may have little realized impact on resulting HC estimates. Iwasaki et al. (2015) compared SSDs constructed using NOECs to those based on EC_{10} s and found that point estimates of HC_5 s based on EC_{10} s were 1.2 (range of 0.6–1.9) times higher. It will invariably be necessary to accommodate a range of metrics for assessing toxicity in SSDs, and studies like that undertaken by Iwasaki et al. (2015) provide reassurance that, on average, the choice of metric is potentially not critical. However, this average outcome

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

bias that depends on toxicity metrics can have ramifications for individual cases in environmental regulation. For example, when HCx values from SSDs are used to manage wastewater discharges, with permits awarded based on the assumption that the discharge is 'safe' by a certain number of dilutions required to achieve the HCx. While a bias of 1.2 may seem small, in a permitting situation if you require 1200 instead of 1000 dilutions, the permit conditions are now being exceeded. If we apply the maximum bias of 1.9 overserved across the studies examined by Iwasaki et al. (2015), the conditions are now being exceeded by nearly double. In that setting the choice of toxicity metric can be of great importance, and a clear framework outlining the required methods to be adopted is essential.

As discussed in (Fisher and Fox 2023) the derivation of NSEC is critically dependent on the estimation of uncertainty in the α (y-intercept) parameter. This means that the statistical methods used must be appropriate and able to accurately capture the level of variability observed in the control treatment. Furthermore, the NSEC will be sensitive to sample size in a similar way to the NOEC (Fisher and Fox 2023). This is verified by our simulation study, which clearly shows that for data based on a sigmoidal model the NSEC declines (becomes more conservative) as sample size increases. This is expected, because experiments with less replication result in greater uncertainty in parameter estimation, including α , which reduces the lower-bound estimate and results in higher estimates of the NSEC value. The impact of experimental design on the estimation of NOEC is well understood (Green et al. 2018) and is directly analogous to the NSEC. Guidance on appropriate experimental design and statistical methods for the estimation of NSEC toxicity values are clearly warranted if these approaches are to be adopted (Fisher

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

and Fox 2023). In addition, it may be reasonable to adopt a similar approach to that of (Negri et al. 2021) where the estimated NSEC values are compared against a relevant EC_x level, and the lower of the two adopted to ensure an appropriate level of protection is maintained, depending on the context of the analysis being undertaken.

While the N(S)EC estimates showed some dependence on sample size, as well as the number of treatment concentrations, values were lower than the EC₁₀ for even the most poorly designed experiment in our simulation study, suggesting that experimental designs would have to be extremely poor for the NSEC to become a less conservative estimate of toxicity than the commonly used EC₁₀ values, and is therefore more appropriate as a measure of no-effect. An interesting result of the simulation study (where true EC_x values are known) was the fact that estimates of EC_x can also be influenced by sampling effort, with some scenarios resulting in substantial overestimates of EC_x values for poor experimental designs. Bias in the estimation of EC_x highlights that issues associated with appropriate experimental design are not specific to estimation of NSEC and should be considered carefully by ecotoxicologists more broadly. The logistics and cost associated with very large experimental designs may put them out of reach for common regulatory studies. However, many designs that we see in current commercial testing laboratories favour replication within treatments, likely because of the requirement of the one-way ANOVA methods used to generate NOECs. We have found that re-distributing replication within treatments across a greater number of treatments can often be achieved with very little additional cost.

This article is protected by copyright. All rights reserved.

CONCLUSIONS

Overall we have demonstrated that the NSEC method proposed by (Fisher and Fox 2023) has the potential to provide an effective estimate of “No-Effect” Concentration that can be used when response data do not show a clear threshold. Embedded within a model averaging approach, the NSEC and NEC can be combined to yield estimates of no-effect concentrations, along with estimation of their uncertainty within a single analysis framework. The outcome is a framework for CR analysis that is robust to uncertainty in the appropriate model formulation, and for which resulting no-effect toxicity estimates can be confidently integrated into the relevant risk assessment framework, such as the Species Sensitivity Distribution (SSD).

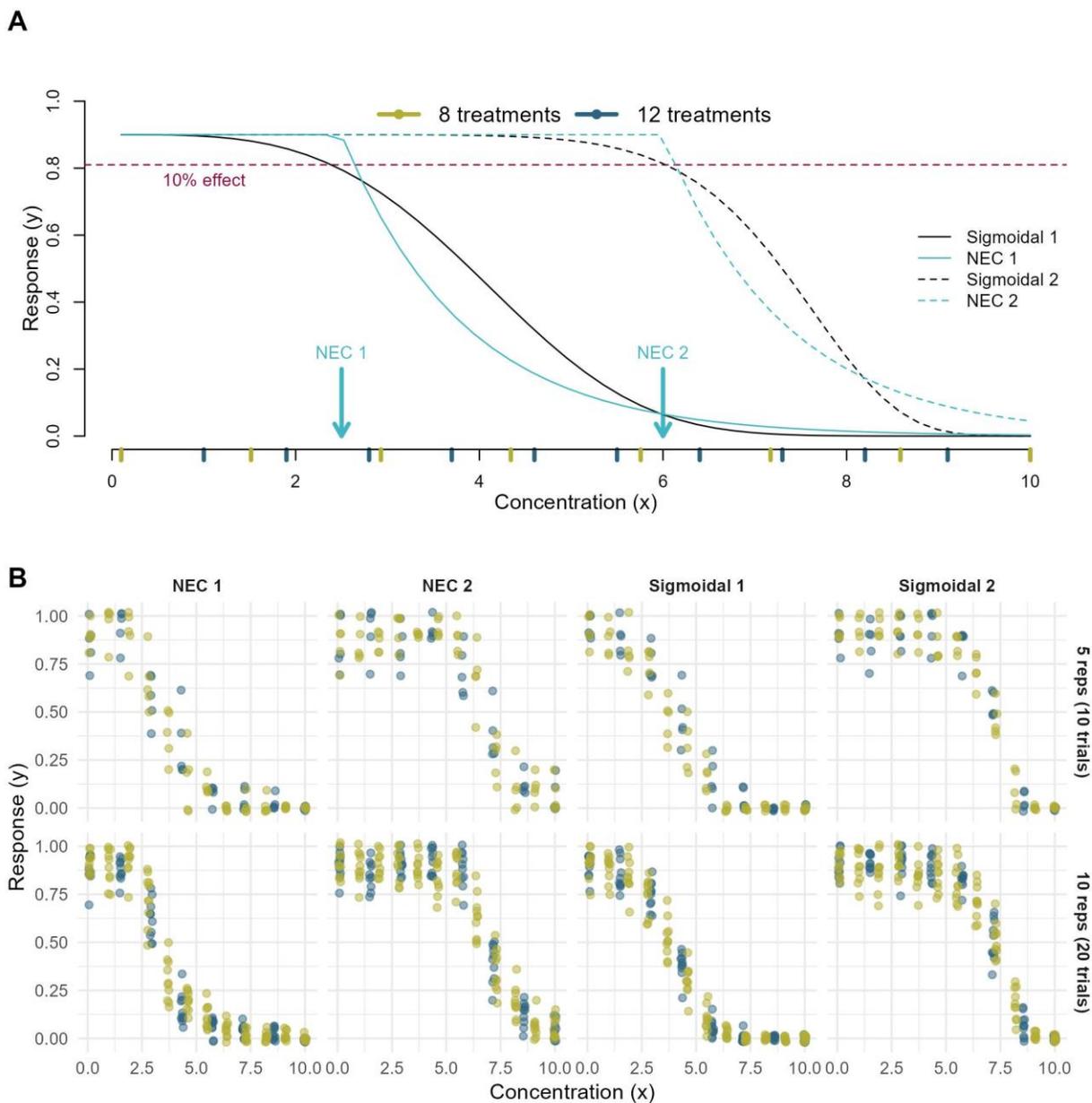


Figure 1. The four CR curves generated for the simulation study, including two NEC and two sigmoidal curves (based on equations 1 and 2 respectively). A) Theoretical curves for all four scenarios. For all scenarios the y-intercept parameter (mean value of the response for the control) was set at 0.9, and for the two NEC models we used the same exponential decay rate of 0.75. The dashed dark red line indicates the theoretical EC10, and the downward blue arrows the position of the NEC for each of the NEC scenarios. Note that

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

there is no theoretical NEC for smooth sigmoidal curves. Coloured tick marks along the concentration (x) axis show the placement of the simulated treatment locations for the 8 treatment (green) and 12 treatment (dark blue) designs. B) Plotted realizations for each scenario for all four scenarios, showing an example dataset for a design with 8 treatments (dark blue circles) and a design with 12 treatments (green circles). The upper row shows data simulated using the least replicated design (5 replicates with 10 trials each) and the lower row shows data for the most replicated design (10 replicates with 20 trials each).

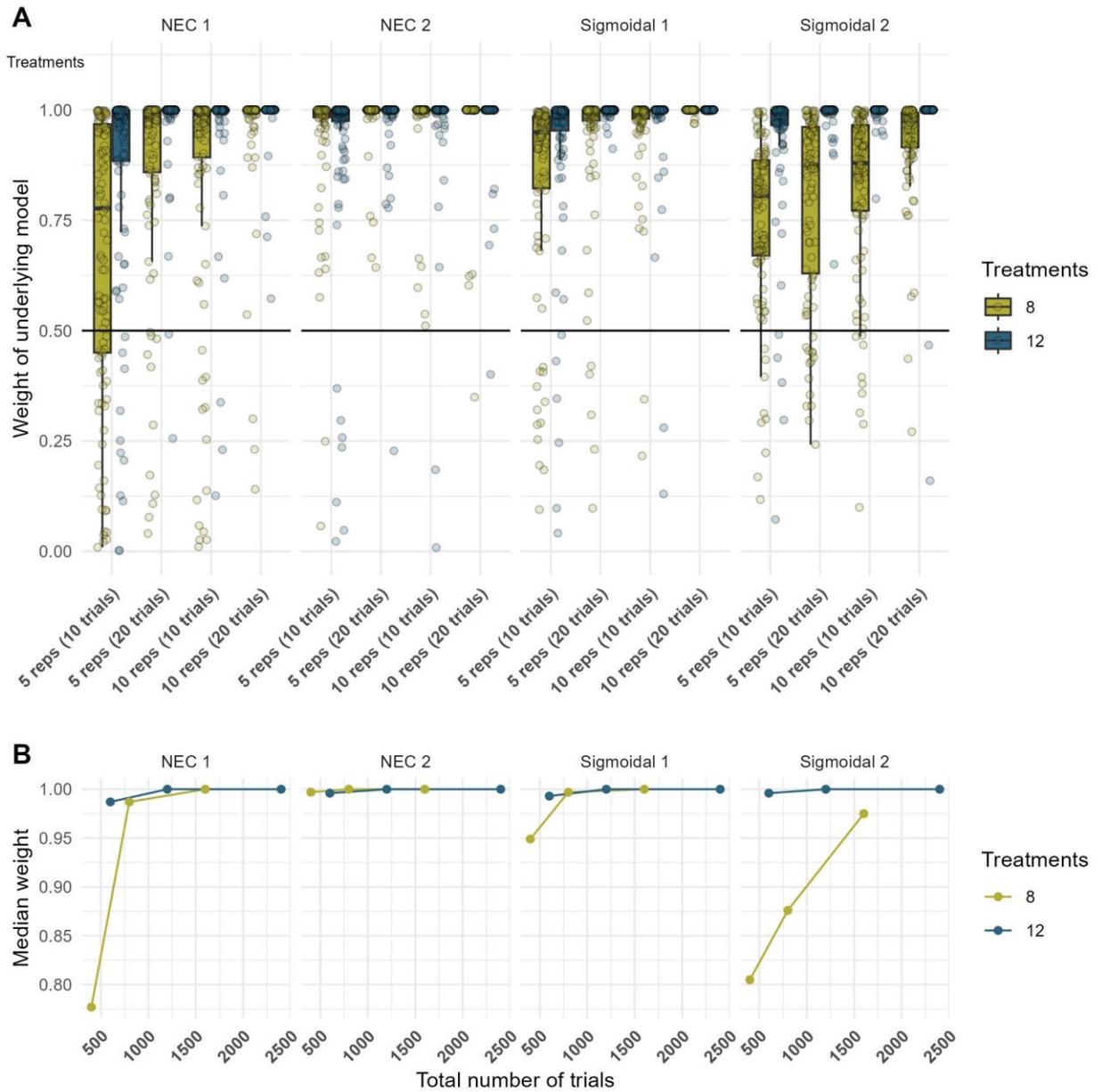


Figure 2. DIC-based model weight for the underlying data generating model (equation 1 or equation 2) relative to the alternative model when fitted using jagsNEC (Fisher et al. 2020) to data simulated according to four theoretical curves, including two NEC threshold models (NEC 1 and NEC 2, equation 1) and two sigmoidal models (Sigmoidal 1 and Sigmoidal 2, equation 2, see Table 2 for details). Simulations were run for a range of sampling designs, including 8 or 12 treatment levels, 5 or 10 replicates within each

This article is protected by copyright. All rights reserved.

Accepted Article

treatment, and 10 or 20 binomial trials (individual survival estimates). The upper plots (A) show box plots based on all simulation outcomes, whereas the lower plots (B) show the median weight across simulations as a function of the total number of replicates.

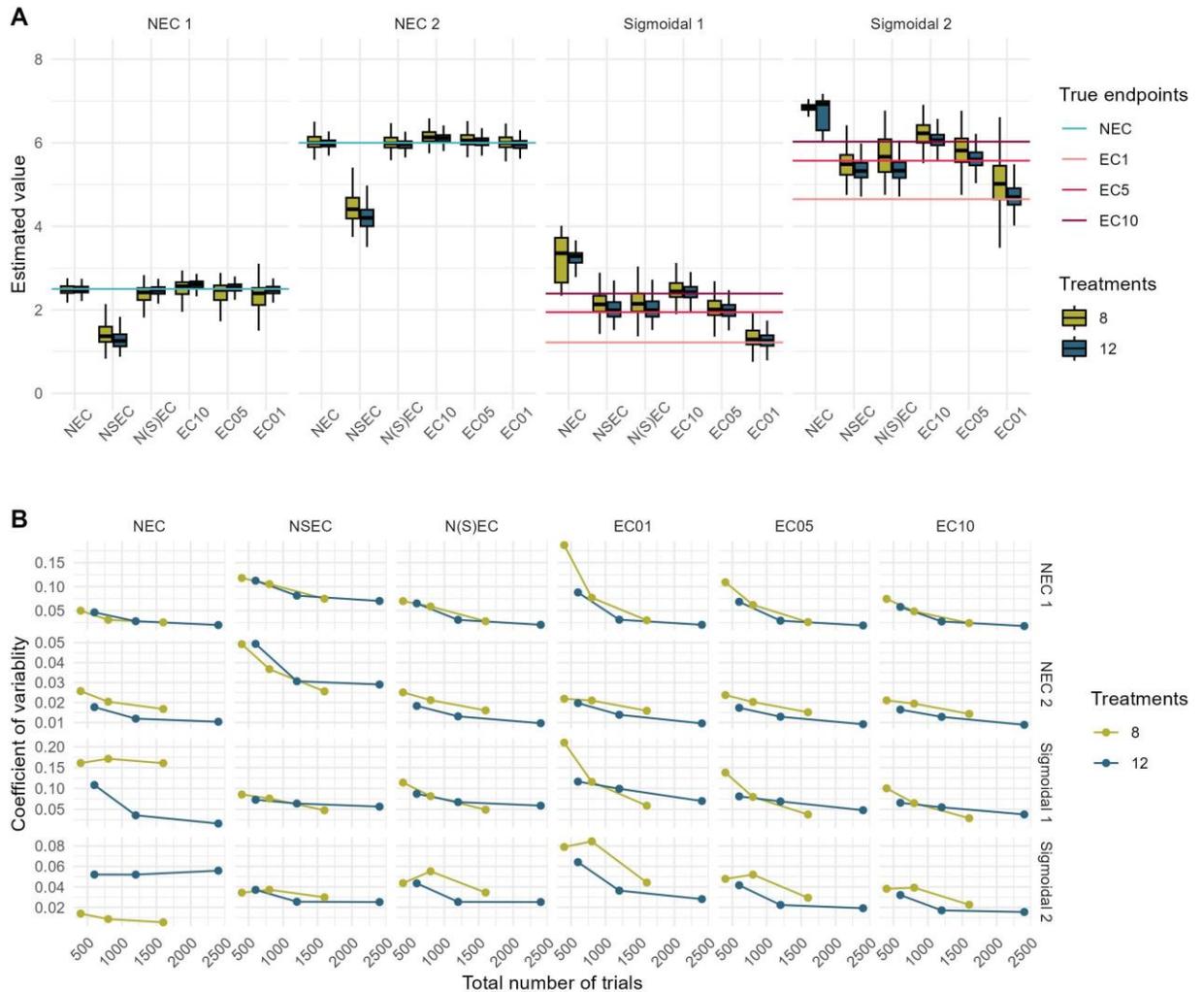


Figure 3. Toxicity estimates (A) and their coefficient of variability (B) from data simulated according to four theoretical curves, including two NEC threshold models (NEC 1 and NEC 2, equation 1) and two sigmoidal models (Sigmoidal 1 and Sigmoidal 2, equation 2, see Table 2 for details). Simulations were run for a range of sampling designs, including 8 or 12 treatment levels, 5 or 10 replicates within each treatment, and

Revised manuscript 10th July 2023

10 or 20 binomial trials (individual survival estimates). All values were estimated using jagsNEC (Fisher et al. 2020) The estimated NSEC value is based on a 99% certainty level, and N(S)EC and all EC_x estimates are a DIC weighted average of MCMC based model fits for both the NEC threshold (equation 1) and the sigmoidal (equation 2) models. NEC values are estimated based only on the NEC model fit, and NSEC is based on the sigmoidal model fit. Also shown are the true toxicity estimates for each scenario (coloured horizontal lines). For the NEC scenarios this is the true theoretical NEC used in simulations. For the sigmoidal scenarios there is no theoretical NEC, and theoretical EC₁, EC₅, EC₁₀ are shown instead. The coefficient of variability is calculated according to the method of Lovitt and Holtzclaw and provides a robust measure of coefficient of variation (Arachchige et al. 2022).

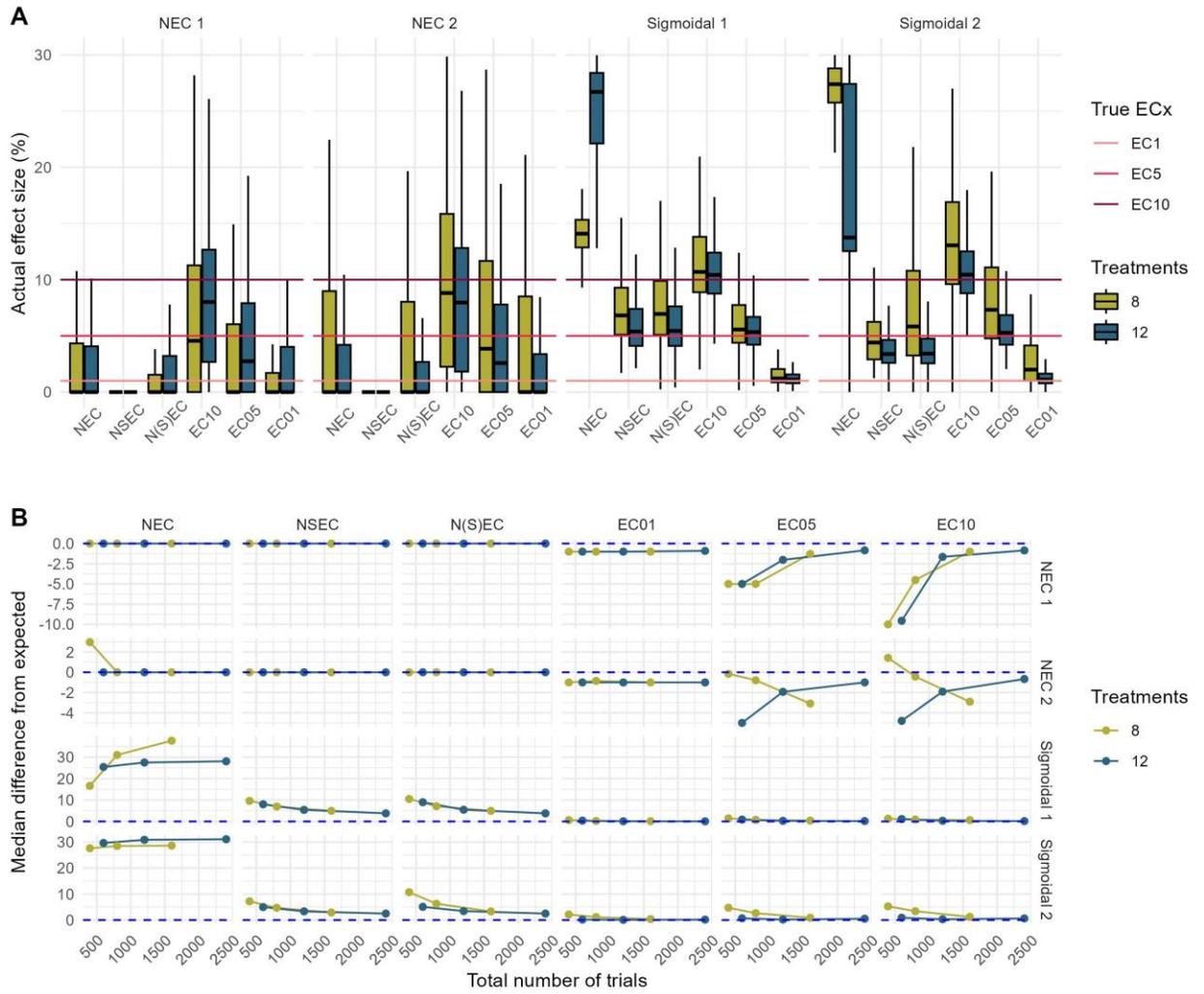


Figure 4. Actual estimated effects of a range of toxicity estimates (A) and their mean difference from expected (B) estimated from data simulated according to four theoretical curves, including two NEC threshold models (NEC 1 and NEC 2, equation 1) and two sigmoidal models (Sigmoidal 1 and Sigmoidal 2, equation 2, see Table 2 for details). Simulations were run for a range of sampling designs, including 8 or 12 treatment levels, 5 or 10 replicates within each treatment, and 10 or 20 binomial trials (individual survival estimates). All values were estimated using jagsNEC (Fisher et al. 2020), as described in

Figure 3. Colored lines show the true EC1, EC5 and EC10 effects. For the NEC, NSEC and N(S)EC the theoretically expected values are 0.

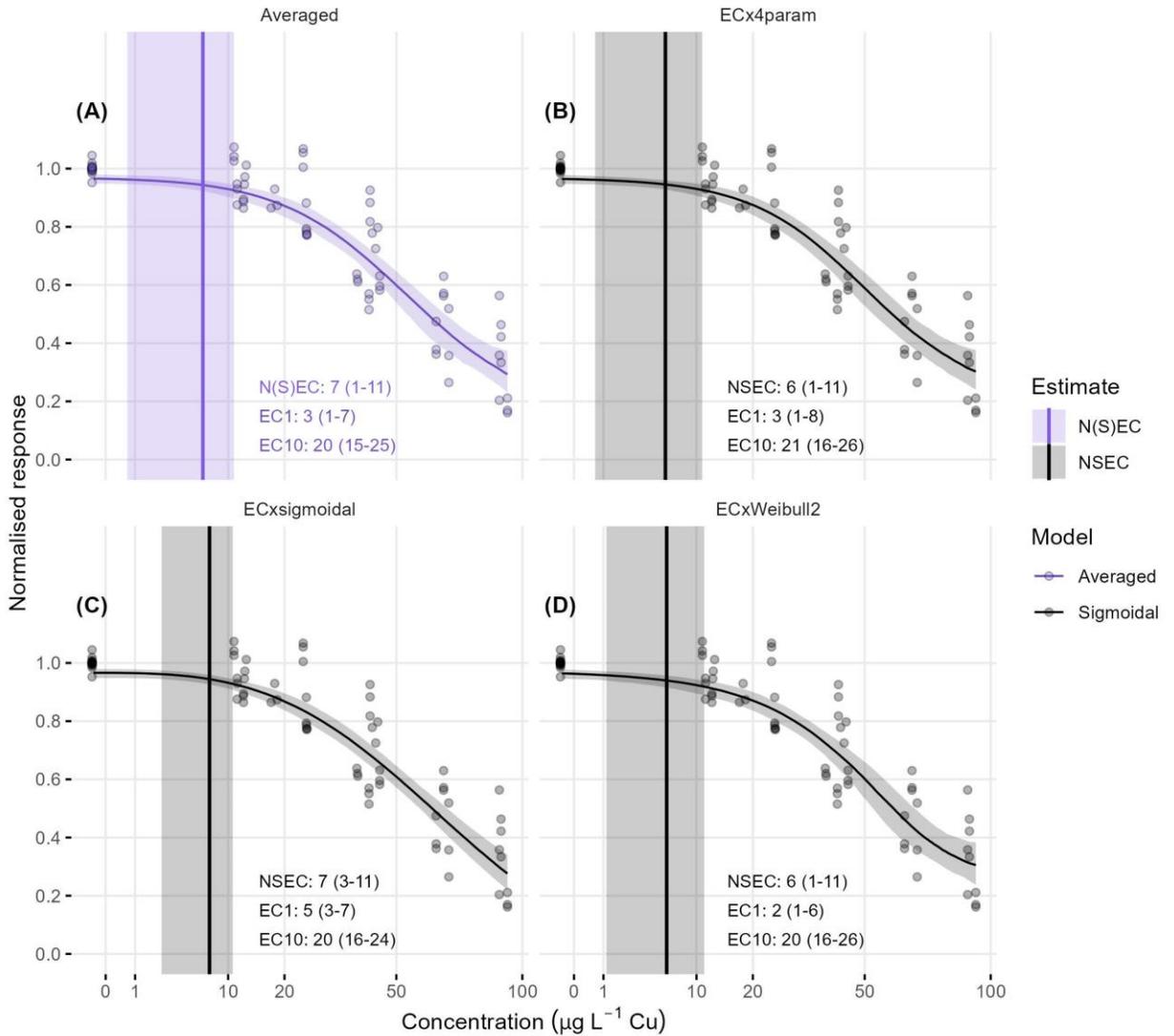


Figure 5. Concentration response relationships for the effects of copper to the Antarctic microalga *Cryothecomonas armigera* (Koppel et al. 2017). The model average (A) and contributing individual models (only those with weights >0, B-D) are shown. The vertical blue lines indicate the estimated NEC value based on the two NEC threshold models (C-D); the vertical black line indicates the NSEC estimated from the Weibull model (B), and

the vertical purple line indicates the model averaged “no effect” concentration (A), denoted as the N(S)EC to indicate it is a weighted average of both NEC and NSEC values.

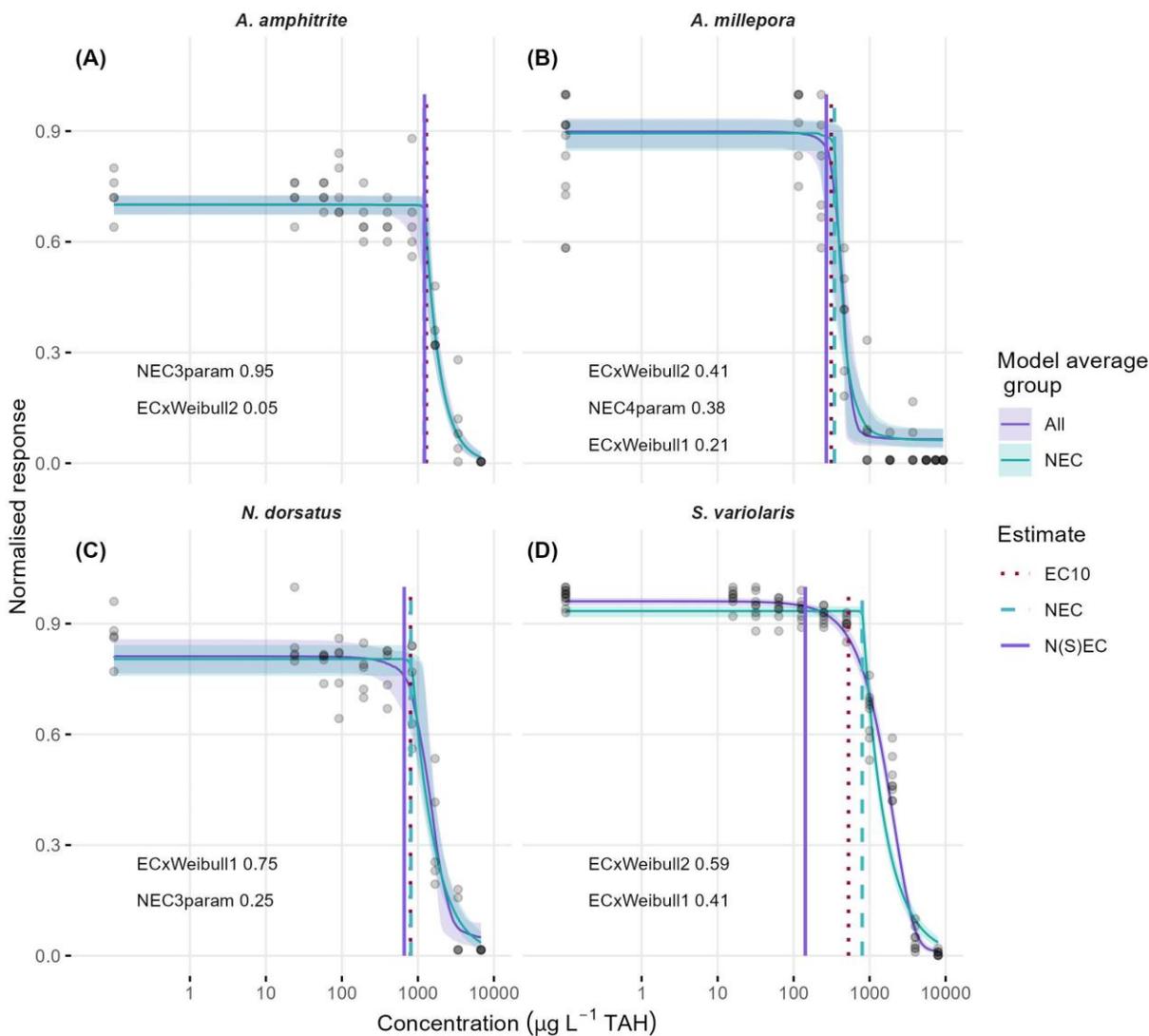


Figure 6. Concentration response relationships for the effects of 40% weathered Ichthys condensate WAF on four tropical marine species. Measured time-weighted average concentrations are expressed as Total Aromatic Hydrocarbons (TAH) on a log scale. Curves are model-averaged Bayesian non-linear beta model fits with 95% credible

Revised manuscript 10th July 2023

intervals indicated by shaded ribbon where all models are included “All” or only NEC models are included “NEC”. Binomial response data are the proportion of successes (*A. millepora*, and *S. variolaris*, B, D); growth rate response data taking values >1 are normalized relative to the maximum value (*N. dorsatus*, C); and growth rate response data taking values <1 are modelled on the original scale (*A. Amphitrite*, A). The vertical lines indicate the estimated NEC, EC10 and model averaged “no effect” N(S)EC values respectively. The model averaged “no effect” is denoted as the N(S)EC to indicate it is a weighted average of both NEC and NSEC values. Note the individual NEC, EC10 and N(S)EC values are very similar and therefore difficult to distinguish for panels A, B and C.

REFERENCES

- Arachchige CN, Prendergast LA, Staudte RG. 2022. Robust analogs to the coefficient of variation. *Journal of Applied Statistics* 49:268-290.
- Bellio R, Jensen J, Seiden P. 2000. Applications of likelihood asymptotics for nonlinear regression in herbicide bioassays. *Biometrics* 56:1204-1212.
- Brooks ME, Kristensen K, Van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Machler M, Bolker BM. 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal* 9:378-400.
- Bürkner P-C. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80:1-28.
- Burnham K, Anderson D. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed.* Springer, New York, NY, USA.
- Chapman PM, Wang F. 2000. Issues in ecological risk assessment of inorganic metals and metalloids. *Human and Ecological Risk Assessment* 6:965-988.
- Chèvre N, Slooten KBv, Tarradellas J, Brazzale AR, Behra R, Guettinger H. 2002. Effects of dinoseb on the life cycle of *Daphnia magna*: Modeling survival time and a

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

proposal for an alternative to the no-observed-effect concentration.
Environmental Toxicology and Chemistry: An International Journal 21:828-833.

de Bruijn JH, Hof M. 1997. How to measure no effect. Part IV: How acceptable is the EC_x from an environmental policy point of view? *Environmetrics: The official journal of the International Environmetrics Society* 8:263-267.

Dormann CF, Calabrese JM, Guillera-Aroita G, Matechou E, Bahn V, Bartoń K, Beale CM, Ciuti S, Elith J, Gerstner K, Guelat J, Keil P, Lahoz-Monfort JJ, Pollock LJ, Reineking B, Roberts DR, Schröder B, Thuiller W, Warton DI, Wintle BA, Wood SN, Wüest RO, Hartig F. 2018. Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs* 88:485-504. DOI: <https://doi.org/10.1002/ecm.1309>.

Fisher R, Barneche D, Ricardo G, Fox D. 2021. bayesnec: A Bayesian No-Effect-Concentration (NEC) Algorithm. R package version 1.0.1. <https://CRAN.R-project.org/package=bayesnec>.

Fisher R, Fox D. 2023. Introducing the no-significant-effect-concentration (NSEC). *Environmental Toxicology and Chemistry*.

Fisher R, Ricardo G, Fox DR. 2020. Bayesian concentration-response modelling using jagsNEC. <https://doi.org/105281/ZENODO3966864>.

Fisher R, Walshe T, Bessell-Browne P, Jones R. 2018. Accounting for environmental uncertainty in the management of dredging impacts using probabilistic dose-response relationships and thresholds. *Journal of Applied Ecology* 55:415-425. DOI: 10.1111/1365-2664.12936.

Fox DR. 2008. NECS, NOECS and the EC {x}. *Australasian Journal of Ecotoxicology* 14:7-9.

Fox DR. 2010. A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology. *Ecotoxicology and environmental safety* 73:123-131. DOI: <http://dx.doi.org/10.1016/j.ecoenv.2009.09.012>.

Fox DR. 2012. Response to landis and chapman (2011). <https://doi.org/101002/ieam1264>.

Fox DR. 2016. Contemporary methods for statistical design and analysis. In Blasco J, Chapman P, Campana O, Hampel M, eds, *Marine Ecotoxicology: Current Knowledge and Future Issues*. Elsevier, San Diego, CA, USA.

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

Fox DR, Landis WG. 2016a. Comment on ET&C perspectives, November 2015—A holistic view. *Environmental Toxicology and Chemistry* 35:1337-1339.

Fox DR, Landis WG. 2016b. Don't be fooled—A no-observed-effect concentration is no substitute for a poor concentration–response experiment. *Environmental Toxicology and Chemistry* 35:2141-2148.

Fox DR, van Dam RA, Fisher R, Batley GE, Tillmanns AR, Thorley J, Schwarz CJ, Spry DJ, McTavish K. 2021. Recent Developments in Species Sensitivity Distribution Modeling. *Environmental Toxicology and Chemistry* 40:293-308.

Gelman A, Hwang J, Vehtari A. 2014. Understanding predictive information criteria for Bayesian models. *Statistics and computing* 24:997-1016.

Green JW, Springer TA, Holbech H. 2018. *Statistical analysis of ecotoxicity studies*. John Wiley & Sons.

Green JW, Springer TA, Staveley JP. 2013. The drive to ban the NOEC/LOEC in favor of ECx is misguided and misinformed. *Integrated environmental assessment and management* 9:12-16.

Hothorn T, Bretz F, Westfall P. 2008. Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 50:346-363.

Iwasaki Y, Kotani K, Kashiwada S, Masunaga S. 2015. Does the Choice of NOEC or EC10 Affect the Hazardous Concentration for 5% of the Species? *Environmental Science & Technology* 49:9326-9330. DOI: 10.1021/acs.est.5b02069.

Koppel DJ, Gissi F, Adams MS, King CK, Jolley DF. 2017. Chronic toxicity of five metals to the polar marine microalga *Cryptomonas armigera*—Application of a new bioassay. *Environmental Pollution* 228:211-221. DOI: 10.1016/j.envpol.2017.05.034.

Krull M. 2020. Comparing statistical analyses to estimate thresholds in ecotoxicology. *PloS one* 15:e0231149.

Mebane CA, Hennessy DP, Dillon FS. 2008. Developing acute-to-chronic toxicity ratios for lead, cadmium, and zinc using rainbow trout, a mayfly, and a midge. *Water, Air, and Soil Pollution* 188:41.

Negri AP, Brinkman DL, Flores F, van Dam J, Luter HM, Thomas MC, Fisher R, Stapp LS, Kurtenbach P, Severati A. 2021. Derivation of toxicity thresholds for gas

This article is protected by copyright. All rights reserved.

Revised manuscript 10th July 2023

condensate oils protective of tropical species using experimental and modelling approaches. *Marine Pollution Bulletin* 172:112899.

Pires AM, Branco JA, Picado A, Mendonça E. 2002. Models for the estimation of a 'no effect concentration'. *Environmetrics: The official journal of the International Environmetrics Society* 13:15-27.

Posthuma L, Suter II GW, Traas TP. 2001. *Species sensitivity distributions in ecotoxicology. 1st Edition*. CRC press, Boca Raton. 616 pp.

Ritz C, Baty F, Streibig JC, Gerhard D. 2015. Dose-response analysis using R. *PloS one* 10:e0146021.

Spiegelhalter DJ, Best NG, Carlin BP, and van der Linde A. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society series B* 64:583–639.

Van Der Hoeven N. 1997. How to measure no effect. Part III: statistical aspects of NOEC, ECx and NEC estimates. *Environmetrics: The official journal of the International Environmetrics Society* 8:255-261.

Van Straalen NM. 1997. How to measure no effect. Part II: Threshold effects in ecotoxicology. *Environmetrics: The official journal of the International Environmetrics Society* 8:249-253.

Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner P, Paananen T, Gelman A. 2020. loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian models. R package version 2.3.1, URL <https://CRAN.R-project.org/package=loo>.

Vehtari A, Gelman A, Gabry J. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing* 27:1413-1432.

Warne M, Batley G, van Dam R, Chapman J, Fox D, Hickey C, Stauber J. 2015. Revised Method for Deriving Australian and New Zealand Water Quality Guideline Values for Toxicants. Prepared for the Council of Australian Government's Standing Council on Environment and Water (SCEW). Department of Science. *Information Technology and Innovation, Brisbane, Queensland*.

Warne M, Batley G, van Dam R, Chapman J, Fox D, Hickey C, Stauber J. 2018. Revised Method for Deriving Australian and New Zealand Water Quality Guideline Values for Toxicants – update of 2015 version. *Prepared for the revision of the Australian and New Zealand Guidelines for Fresh and Marine Water Quality*

This article is protected by copyright. All rights reserved.

Australian and New Zealand Governments and Australian state and territory governments, Canberra (48 pp).

Warne MSJ, Van Dam R. 2008. NOEC and LOEC data should no longer be generated or used. *Australasian Journal of Ecotoxicology* 14:1-5.

Table 1. Toxicity estimates currently used for estimating no and low- effects for using in SSD modelling. Definitions are adapted from Warne et al. 2015, with the exception of NSEC which is described in Fisher and Fox 2023.

Toxicity estimate	Definition	Statistical method
NSEC	No significant effect concentration - The concentration at which the mean response is statistically indistinguishable from the mean control response	Interpolation from a CR model (Fisher and Fox 2023)
NEC	No effect concentration - The minimum concentration above which an effect is predicted to occur	Parameter estimate of a CR threshold model (van der Hoeven et al. 1997, Fox 2010, Pires et al. 2002)
NOEC	No observable effect concentration - The highest tested concentration at which the mean response is statistically indistinguishable from the mean control response	Dunnett's test (based on ANOVA)
ECx/ICx/LCx	x% effect/inhibition/lethal concentration - The concentration that is expected to cause a specified effect in x% of a group of organisms or x% effect (ECx); an x% reduction in a non-quantal measurement such as fecundity or growth (ICx); or be lethal to x% of a group of organisms (LCx)	Interpolation from a CR model

Table 2. Scenario parameters used in simulations.

scenarios	α	β	γ	δ (d)	N treatments
Sigmoidal 1	0.9	5.0e-03	-	3.5	8
Sigmoidal 1	0.9	5.0e-03	-	3.5	12
NEC 1	0.9	7.5e-01	2.5	-	8
NEC 1	0.9	7.5e-01	2.5	-	12
Sigmoidal 2	0.9	1.0e-08	-	9.0	8
Sigmoidal 2	0.9	1.0e-08	-	9.0	12
NEC 2	0.9	7.5e-01	6.0	-	8
NEC 2	0.9	7.5e-01	6.0	-	12

Table 3. Successfully fitted models and their associated DIC and DIC based model weights for the concentration response relationships of the effects of copper to the Antarctic microalga *Cryothecomonas armigera*. See the Supplementary information for details of model formulations for each model.

Model	DIC	ω_i
NEC3param	-174.311	0.001
NEC4param	-170.562	0
ECxLinear	-181.141	0.034
ECxExp	-127.252	0
ECxsigmoidal	-185.696	0.334
ECx4param	-185.943	0.377
ECxWeibull1	-183.271	0.099
ECxWeibull2	-184.154	0.154

Table 4. Toxicity estimates based on two case studies, showing NEC, N(S), EC1, EC10 and the estimated % effect value of the reported N(S)EC. Case study one are from effects of copper to the Antarctic microalga *Cryothecomonas armigera* (Koppel et al. 2017), and case study two is the effects of 40% weathered *Ichthys condensate* WAF on four tropical marine species (Negri et al. 2021). All values were estimated using jagsNEC using the default settings (Fisher et al. 2020). N(S)EC and all ECx estimates are a DIC weighted average of MCMC based model fits for both the NEC threshold and the sigmoidal models (see the Supporting Information for more details of individual models). NEC values are estimated based only on the NEC model fits. NOEC values are based on fitting the concentration data as a treatment factor via the package glmmTMB (Brooks et al. 2017) and doing a Dunnett's test via glht from the package multcomp (Hothorn et al. 2008).

Case study	Species	NOEC/LOEC	N(S)EC	NEC	EC1	EC10	N(S)EC % effect
1	<i>C. armigera</i>	11.1/14.2*	7 (1-11)	20.9 (14.9-23.5)	3 (1-7)	20 (15-25)	2.3
2	<i>A. amphitrite</i>	837/ 1684	1220 (500-1360)	1200 (1000-1400)	1280 (732-1440)	1306 (796 – 1490)	2.3
	<i>A. millepora</i>	231/463	270 (25-450)	340 (210-450)	236 (68-451)	325 (201-465)	5.5
	<i>N. dorsatus</i>	837/1684	660 (350-1140)	840 (690-1200)	342 (70-1110)	771 (374-1306)	6.6
	<i>S. variolaris</i>	31.8/62.7	140 (60-200)	800 (710-850)	114 (79-170)	530 (438 – 625)	1.5

Revised manuscript 10th July 2023

* As the concentration data for this case study were individually measured they were grouped using a factor of the square-root transformed concentrations rounded to one significant figure to enable a Dunnett's test to be performed. The reported value represents the cell mean for the lowest non-significant grouping.