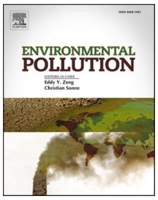
Environmental Pollution 338 (2023) 122708

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol



Improving toxicity prediction of metal-contaminated sediments by incorporating sediment properties☆

Yanfeng Zhang[a,b], Minwei Xie[c,*], David M. Spadaro[a], Stuart L. Simpson[a,d]

[a] *CSIRO Environment, Lucas Heights, NSW, 2234, Australia*
[b] *Tianjin Key Laboratory of Remediation & Pollution Control for Urban Ecological Environment, Key Laboratory of Pollution Processes and Environmental Criteria of Ministry of Education, College of Environmental Science and Engineering, Nankai University, Tianjin, 300350, China*
[c] *State Key Laboratory of Marine Environmental Science, Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, College of the Environment and Ecology, Xiamen University, Xiamen, Fujian, 361102, China*
[d] *Hong Kong Branch of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China*

ARTICLE INFO

Keywords:
Risk assessment
Sediment quality
Reproductive toxicity
Multiple linear regression
Amphipod

ABSTRACT

For the purpose of sediment quality assessment, the prediction of toxicity risk-levels for aquatic organisms based on simple environmental measurements is desirable. One commonly used approach is the comparison of total contaminant concentrations with corresponding water and sediment quality guideline values, serving as a Line of Evidence (LoE) based on chemistry-toxicity effects relationships. However, the accuracy of toxicity predictions can be improved by considering the factors that modify contaminant bioavailability. In this study we used paired chemistry-ecotoxicity data sets for sediments to evaluate the improvement in toxicity risk predictions using bioavailability-modified guidelines. The sediments were predominantly contaminated with metals, and measurements of sediment particle size, total organic carbon (TOC) and acid volatile sulfide (AVS) were used to modify hazard quotients (HQ). To further assess the predictive efficacy of the bioavailability-modified guideline models, sediments with differing contamination levels were tested for toxicity to a benthic amphipod's reproduction. To account for differences between laboratory exposure and field exposure scenarios, where the latter creates greater dilution, both static-renewal and flow-through test procedures were employed, and flow-through resulted in lower dissolved metal concentrations in the overlying waters. We also investigated how lower AVS concentration by oxidation modified the toxicity. This study reaffirmed that consideration of factors that influence contaminant bioavailability improves toxicity risk predictions, however the improvements may be modest. The sediment particle size data had the greatest influence on the modified HQ, indicating that higher percentage of fine particle size (<63 μm) contributed most to a lower predicted toxicity. The comparison of the static-renewal and flow-through test results continue to raise important questions about the relevance of static or static-renewal toxicity test results for risk assessment decisions, as both these test designs may cause unrealistically high contributions of dissolved metals in overlying waters to toxicity. Overall, this study underscores the value of incorporating outcomes from simple and routine sediment analysis (e.g., particle size, TOC, and consideration of AVS) to enhance the predictive efficacy of toxicity risk assessments in the context of sediment quality risk assessment.

1. Introduction

The use of multiple lines of evidence (LoE) is recommended to improve the robustness of sediment quality assessments (Bay and Weisberg, 2012; Simpson and Batley, 2016). Common LoE include measurements of chemistry (contaminant concentrations), ecotoxicology (lethal and sublethal effects), bioaccumulation (uptake of chemicals) and ecology (benthic community diversity and populations). However, the collection of multiple LoE can be prohibitively expensive for small-scale assessments, and this makes it more desirable to be able

---

☆ This paper has been recommended for acceptance by Sarah Harmon.

* Corresponding author.

E-mail address: minweixie@xmu.edu.cn (M. Xie).

https://doi.org/10.1016/j.envpol.2023.122708
Received 22 June 2023; Received in revised form 28 August 2023; Accepted 5 October 2023

to predict risks of adverse effect of contaminants using routine and cost-effective measurements.

The prediction of the risk of sediment toxicity based solely on comparison of total concentrations of major contaminants with sediment-quality guideline values (SQGVs) is often insufficient. Incorporating factors that modify contaminant bioavailability into models can improve toxicity predictions (Di Toro et al., 2005; Maruya et al., 2012; Simpson and Batley, 2007). For metal contaminants, measurements of sediment particle size, and concentrations of major contaminant binding phases, such as organic matter and acid volatile sulfide (AVS) have been demonstrated to improve prediction of contaminant bioavailability and toxicity risks (Campana et al., 2013; Di Toro et al., 2005; Schlekat et al., 2016).

When using laboratory-generated contaminant exposure and effects data to predict ecotoxicity risks posed by contaminated sediments in the field, it is also important to understand how differences between lab and field conditions influence contaminant exposure routes and the risk prediction (Burton et al., 2005; Harrison et al., 2019; Mann et al., 2010). Contaminant binding is modified when sediments are disturbed by the bioturbation activities of organisms within their local environment or by hydrodynamic forces from water currents (Amato et al., 2016; Ciutat and Boudou, 2003; Xie et al., 2019).

In this study we utilised a paired chemistry and ecotoxicity data set (N = 65) to explore simple models to predict toxicity risks. Firstly, we considered major contaminant concentrations in the form of a hazard quotient (HQ), and secondly, we incorporated sediment properties measurements to modify the HQ. We then expanded this data set with new paired data (N = 26) from laboratory toxicity tests with the epibenthic deposit-feeding estuarine-marine amphipod *Melita plumulosa*. These tests encompassed diverse sediment samples, including 19 estuarine sediments collected from the Sydney region, along with an additional 7 sediments derived from the initial 19 sediments subjected to AVS manipulations. These latter 7 sediments underwent modifications in AVS through resuspension and oxidation, resulting in varying reductions in the concentration of this strong metal binding phase. Both static-renewal and flow-through test procedures were used, intended to result in differing contaminant concentrations in overlying water. The hypotheses evaluated were that (i) a HQ modified using sediment particle size, total organic carbon (TOC) and AVS as factors that influence metal bioavailability (referred to as HQ*) would improve in toxicity risk predictions, and (ii) an improved prediction of toxicity risk-levels would be observed across static-renewal and flow-through data sets, particularly in sediments where AVS concentrations have been reduced through oxidation.

2. Materials and methods

2.1. Historical sediment chemistry and toxicity data sets

Sediment data sets where metals were the dominant contaminants were selected from past sediment toxicity studies (Amato et al., 2014; Simpson et al., 2014; Simpson et al., 2020; Simpson et al., 2013). The paired chemistry and toxicity data for these 65 sediments have been consolidated and summarised in Table 1, while specific metal concentrations are detailed in Table S1 within the Supplementary Material. Sediment chemistry measurements encompassed total concentrations of the metals (Cd, Cr, Cu, Ni, Pb and Zn, and the metalloid As), along with additional parameters including fine particle size fraction (f_{FP}, % <63 μm), total organic carbon (% TOC) and acid-volatile sulfide (AVS) concentration (μmol/g).

2.2. New sediments collection and analyses

A total of 19 surficial sediments (0–10 cm depth) were collected from estuarine locations in the Sydney region (New South Wales, Australia). These sediments exhibited relatively low levels of organic contaminants but varying levels of heavy metal contamination and physicochemical properties (Xie et al., 2021; Zhang et al., 2020). For 7 of the sediments, replicate subsamples were used in toxicity tests following a period of oxidation that occurred by resuspending the sediment in seawater, and was undertaken to provide sediments with the same metal concentration but lower AVS concentration. This resulted in 26 new sediments for tests. The methods used for sediment collection, manipulation and analyses are described in Section S1 of the Supplementary Material.

The methods for analyses of metal concentrations (sediment and dissolved), f_{FP}, TOC and AVS are provided in the Supplementary Material. In brief, particle size distribution was measured with Malvern Mastersizer 3000 (Malvern Panalytical) and the fraction <63 μm used to define f_{FP}. TOC was determined using the method of high temperature combustion coupled infrared detection with a Shimadzu TOC analyser. AVS was analysed with a rapid method for measuring the concentration of sulfides released from the sediment by dilute acid. Analyses of total recoverable metals (TRM) on bulk sediment samples were made following low pressure aqua regia digestion of sediment in a microwave digestion system (MARS 5, CEM). Overlying water from toxicity tests were filtered (0.45 μm) and acidified with concentrated HNO_3 (0.2%, v/v). Metal concentrations of As, Cd, Cr, Cu, Ni, Pb and Zn in water and acid digests were determined using inductively coupled plasma – atomic emission spectrometry (Agilent 720 ICP-OES). With the exception of polycyclic aromatic hydrocarbons (PAHs) and total recoverable hydrocarbons (TRHs), other organic contaminants (e.g., pesticides, polychlorinated biphenyls) are regarded as negligible based on past studies within the sediment source areas (Birch, 2017; Chariton et al., 2010; Dafforn et al., 2012). Organic contaminants included PAHs and TRHs were analyzed to confirm that metals are the dominant contaminants in the tested sediments, and any sediments with concentrations of PAHs and TRHs of concern were excluded from this study (Simpson et al., 2020).

2.3. Sediment toxicity tests

The epibenthic amphipods, *Melita plumulosa,* were obtained from laboratory cultures. Toxic effects of exposure to whole-sediment to the survival and reproduction of *M. plumulosa* were assessed over a period of 10 days. Clean seawater used in the tests was from the South coast of Sydney, Australia. Two test methods were used that created different overlying water environments: a static-renewal test (Spadaro and Simpson, 2016) and a flow-through test (Zhang et al., 2020), where the flow-through method resulted in lower concentrations of released metals in the overlying waters. Four replicates were completed for every test. Both methods are described in detail in the Supplementary Material.

Briefly, at the start of the tests, six gravid females (gravid for <24 h) and six males were randomly assigned to each test chamber. All treatments (static-renewal and flow-through) were fed at a rate of 0.5 mg food per amphipod on day 0 and 5. At the end of the tests, the survival rate of the adult amphipod was recorded. Subsequently, the count of embryos was conducted under microscopy. Juvenile amphipods were also meticulously collected by sieving the sediment, and their numbers were counted. The count of offspring (including embryos and juveniles) per female amphipod was then recorded as the reproduction endpoint. Both the survival rate and the reproduction endpoint were expressed as percentages of the control group. To ensure that physicochemical parameters remained within acceptable ranges, measurements of the overlying water were undertaken at the beginning and periodically throughout the tests, including dissolved oxygen (>85% saturation), pH (8.0 ± 0.2), salinity (30 ± 2‰), temperature (21 ± 1 °C) and ammonia (<3 mg NH_3–N/L). Toxicity was detected when the reproductive output measured less than 80% of the control and exhibited statistically significant differences ($p < 0.05$) compared to the controls. This assessment classified toxicity into five categories: No toxicity, Low, Moderate, High, and Very high (Table S2, Supplementary Material).

Table 1
Historical sediment data overview: sediment properties, hazard quotients (HQ and HQ*) and toxicity results.

Sediment	f_{FP} (% <63 μm)	TOC (%)	AVS (μmol/g)	HQ	HQ*	Survival (%)		Reproduction (%)	
						Mean	SE	Mean	SE
1	10	0.54	0.6	0.17	0.47	98	7	98	7
2	2.1	0.75	41	0.17	0.68	94	3	67	5
3	20	1.0	1.1	0.23	0.36	98	14	98	14
4	7.2	0.62	4.1	0.25	0.67	88	4	60	10
5	21	0.50	0.1	0.32	0.60	85	3	70	14
6	8.0	0.62	0.5	0.33	1.08	86	9	70	8
7	65	3.0	3.4	0.48	0.27	103	11	97	14
8	3.9	1.8	0.4	0.50	2.13	94	3	53	7
9	5.4	0.78	5.0	0.56	1.63	89	7	74	13
10	32	2.2	0.3	0.60	0.73	102	6	98	6
11	59	0.91	9.4	0.62	0.39	94	2	80	10
12	94	4.4	15	0.78	0.27	93	4	91	19
13	61	1.3	3.2	0.85	0.57	99	2	94	6
14	37	2.2	0.4	0.88	0.95	98	4	89	13
15	44	2.8	12	0.94	0.57	93	4	83	10
16	78	1.5	2.1	0.97	0.58	97	6	87	9
17	38	2.0	6.2	1.07	0.82	104	4	79	16
18	64	3.2	1.5	1.08	0.67	96	4	80	9
19	46	3.0	9.2	1.24	0.75	105	4	67	16
20	48	5.2	1.8	1.29	0.85	99	4	85	14
21	84	7.4	22	1.39	0.45	93	8	75	20
22	53	5.2	17	1.40	0.66	100	8	83	12
23	79	4.6	58	1.52	0.49	108	4	91	7
24	80	4.7	3.3	1.56	0.71	94	3	94	9
25	92	4.3	5.4	1.58	0.64	86	2	75	6
26	68	4.3	23	1.66	0.66	84	10	87	10
27	95	2.7	4.5	1.70	0.74	101	2	85	7
28	25	1.4	1.9	1.73	2.07	52	4	22	18
29	93	3.4	15	1.76	0.65	89	2	80	18
30	87	3.2	9.8	1.89	0.77	98	4	87	15
31	97	2.8	18	2.06	0.74	95	4	79	7
32	37	2.2	0.4	2.06	2.23	100	2	47	2
33	15	4.0	7.4	2.33	2.71	77	10	23	17
34	50	3.9	9.0	2.41	1.33	102	2	92	7
35	65	4.8	42	2.48	0.93	98	4	56	10
36	65	4.5	13	2.66	1.17	90	11	94	23
37	37	2.2	0.4	2.67	2.88	104	4	25	4
38	77	4.3	10	2.75	1.13	113	7	71	3
39	88	5.3	13	2.76	0.98	103	3	93	11
40	25	3.0	1.2	2.91	3.28	63	8	4	4
41	59	4.5	0.5	3.12	2.20	78	3	26	7
42	21	8.4	42	3.32	2.22	104	4	65	23
43	86	3.4	22	3.95	1.44	95	3	84	19
44	15	1.9	0.6	4.51	6.23	95	9	57	14
45	53	4.0	2.4	5.33	3.43	105	5	61	6
46	65	4.1	13	5.67	2.54	91	6	65	5
47	84	4.0	35	7.58	2.58	105	6	52	11
48	84	6.8	23	9.90	4.83	43	10	1	1
49	40	4.8	2.0	0.24	0.18	100	0	100	13
50	20	0.70	0.5	0.26	0.47	90	4	100	5
51	98	4.5	0.4	0.36	0.19	100	3	100	13
52	60	3.5	0.4	0.37	0.28	100	4	100	1
53	49	2.4	2.0	0.39	0.28	102	7	103	11
54	12	1.4	40	0.41	0.53	100	2	100	8
55	50	1.7	0.5	0.46	0.43	90	8	100	5
56	70	2.6	0.5	0.53	0.37	90	6	100	5
57	98	4.5	5.0	0.66	0.26	105	7	112	7
58	98	4.5	5.0	0.66	0.26	100	4	100	6
59	90	3.1	0.5	0.73	0.43	85	2	100	5
60	100	3.6	0.5	0.75	0.40	83	5	100	5
61	65	2.2	11	0.81	0.41	102	4	109	35
62	95	1.6	15	1.21	0.49	100	0	118	13
63	98	2.9	2.1	1.27	0.59	90	4	109	8
64	98	3.3	18	1.37	0.48	98	6	112	14
65	25	1.8	5.1	3.87	3.96	25	2	0	0

f_{FP}: fraction of fine particles. TOC: total organic carbon. AVS: acid-volatile sulfide. HQ: hazardous quotient based on the sediment concentration and sediment quality guideline value (SQGV) of the metals.
HQ* is the hazardous quotient modified using f_{FP}, TOC and AVS, as described by Eq. (6).
Amphipod survival and reproduction data are expressed as % control.
Sediments ordered by increasing HQ, where the table has a split at sediments 48/49 at which point 17 sediments (*italicised numbers*) are then described that had 0% or ≥100% toxicity (% control) and treated differently to the first 48 when constructing the exposure-effect model.

In the static-renewal tests, the organisms were exposed to contaminants in 250 mL glass beakers that contained 40 g of sediment and 200 mL of overlying seawater. Samples of overlying water were collected on the first and the last day of the tests, as well as on Days 3, Day 5 and Day 7 when the overlying water was exchanged with clean seawater. Time-weighted average concentrations were obtained from the average of the five measurements. Sediment renewal was carried out on Day 5. The static-renewal tests were conducted at constant temperature (21 ± 1 °C) in an environmental chamber (Labec Refrigerated Cycling Incubator, Laboratory Equipment) on a 12-h light/12-h dark rotational cycle and aeration was provided.

In the flow-through tests, the organism exposures were conducted in chambers within a tank that allowed for seawater exchange between the chamber and larger tank to create greater dilution of contaminants released from sediment into overlying water than occurred in the static-renewal tests (Figure S1, Supplementary Material). A pump circulated the seawater from the top of the tank and pumped it back into the exposure chambers through the port. Samples of overlying water were collected on the first and the last day of the tests, as well as on Days 3 and Day 7. Time-weighted average concentrations were calculated from the average of the four measurements. The flow-through tests were conducted in an air-conditioned laboratory (21 ± 1 °C) with normal daylight cycles and aeration provided with the water recycling.

2.4. *Hazardous quotient and toxic unit*

Hazard quotients (HQ) for the sediments were calculated based on the contaminants concentrations of the sediment as described previously (Long, 2006; Simpson et al., 2013) using each SQGV (Table S1, Supplementary Material) (ANZG, 2018). Because the sediments were mainly metal-contaminated, HQ was calculated using concentrations of the seven metals (M) of As, Cd, Cu, Cr, Ni, Pb, and Zn, i.e. $\mathrm{HQ} = \sum_{n}([\mathrm{M}]/\mathrm{SQGV})/n$, where [M] is the metal concentration (TRM) in sediment, SQGV is the corresponding sediment quality guideline value, and n is the number of contaminants (i.e., 7 metals in this case).

Toxic units (TU) for overlying waters were calculated as the mean of time-weighted average concentrations of dissolved metals (dM) in the overlying waters divided by their respective marine water quality guideline value (WQGV) (ANZG, 2018; Golding et al., 2022), i.e., $\mathrm{TU} = \sum_{n}(\mathrm{dM}/\mathrm{WQGV})/n$ (Table 2). Values of HQ and TU < 1 indicate a low risk of toxicity, and values > 1 indicate increasing risk and potential magnitude of toxicity as the values increase.

2.5. *Statistics*

The t-tests were performed to assess a significant reduction in amphipod survival or reproduction in test sediments compared to the controls. The data were tested for normality of distribution (Shapiro-Wilk's test) before other statistical analyses. Significance in all statistical tests was set at the $p < 0.05$ level.

Given that amphipod reproduction is a more sensitive endpoint in response to metal contamination and consistently manifested in the majority of tests, logistic regression was used to build the relationship between reproduction and HQ and TU. Multiple linear regression (MLR) was applied to investigate the effects of HQ and the sediment properties (AVS, TOC and f_{FP}) on the reproduction. Significance in all statistical tests was set at the $p < 0.05$ level. All the statistical analyses were carried out using the software Origin 2017.

The sigmoidal relationship between toxicity to amphipod reproduction and HQ (or TU) was established with logistic regression using Eq. (1).

$$y = \frac{100}{1 + \left(\frac{x}{c}\right)^{b}} \tag{1}$$

where y is the survival (%) or reproduction (%), and x is HQ or TU.

Toxic effects to amphipods may occur from metals in overlying water and porewater, as represented here by TU, and from metals that remain associated with the sediments, represented by HQ. Because metal concentrations in overlying water are influenced by sediment properties

Table 2

New sediments data: Properties of the sediments and hazardous quotient and toxic unit in the tests.

Sediment	f_{FP} (% <63 μm)	TOC (%)	AVS (μmol/g)	HQ	HQ*	TU	
						Static-renewal	Flow-through
1	55	4.0	0.5	0.2	0.16	0.4	0.4
2	91	4.5	0.5	0.7	0.41	0.3	0.3
3	81	4.4	0.5	1.1	0.64	0.5	0.3
4	71	2.4	2.0	1.4	0.82	1.3	0.5
5	63	6.6	0.6	1.4	0.90	0.5	0.3
6	38	1.8	0.6	1.6	1.71	1.2	0.3
7 (Ox)	38	1.8	0.5	2.8	1.75	2.1	0.8
8	35	5.5	12	1.8	1.10	0.3	0.2
9 (Ox)	35	5.5	11	1.8	1.11	0.8	0.3
10	49	7.7	0.7	1.8	1.24	0.7	0.2
11	58	3.7	4.8	1.9	1.03	3.8	0.3
12	61	3.7	13	2.0	0.91	1.6	0.5
13	35	8.7	0.7	2.2	1.77	0.7	0.3
14	33	3.0	18	2.2	1.53	0.3	0.3
15 (Ox)	33	3.0	6.9	2.2	1.72	2.5	0.3
16	36	1.5	0.5	2.5	2.89	1.8	0.8
17 (Ox)	36	1.5	0.5	2.5	2.89	3.4	0.9
18	39	1.8	0.6	1.6	3.85	1.4	0.5
19	40	1.0	0.5	3.2	3.66	0.9	0.3
20 (Ox)	40	1.0	0.6	3.2	3.58	0.3	0.3
21	32	18	56	3.8	1.68	0.3	0.3
22 (Ox)	32	18	13	3.8	2.02	0.4	0.5
23	68	14	0.8	6.7	2.56	3.4	0.3
24	56	6.4	0.5	5.1	3.47	1.9	0.5
25	47	6.9	9.2	6.4	3.36	1.2	1.4
26 (Ox)	47	6.9	0.5	6.4	4.85	3.9	1.0

Seven sediments with (Ox) indicates those that were resuspension-oxidation treated, with the corresponding non-treated sediment being the number above. f_{FP}, TOC, AVS and HQ are defined in Table 1. HQ* is the hazardous quotient modified using f_{FP}, TOC and AVS, is described by Eq. (6).
TU: toxic unit based on the overlying water concentration and water quality guideline value (WQGV) of the metals.

that influence metal binding and hence bioavailability, we sort to describe the influence of AVS, TOC and f_{FP} on the bioavailable metal exposure through the derivation of HQ* that represented the sediment property-modified HQ. HQ* was derived using Eq. (2).

$$HQ^* = HQ \bullet \left(\frac{AVS}{AVS_{Ref}}\right)^{s_1} \bullet \left(\frac{TOC}{TOC_{Ref}}\right)^{s_2} \bullet \left(\frac{f_{FP}}{f_{FP\,Ref}}\right)^{s_3} \tag{2}$$

where AVS_{Ref}, TOC_{Ref} and $f_{FP\text{-}Ref}$ are the reference values (chosen arbitrarily) for AVS (0.5 μmol/g), TOC (1%) and f_{FP} (50%) and s_1, s_2 and s_3 are the slopes, respectively. The relationship between toxicity and the sediment metals modified by these sediment parameters was derived using Eq. (3).

$$\log\frac{100-y}{y} = b \bullet \log HQ + b \bullet s_1 \bullet \log AVS + b \bullet s_2 \bullet \log TOC + b \bullet s_3 \bullet \log f_{FP} + b_0 \tag{3}$$

3. Results and discussion

3.1. Toxicity models derived using historical chemistry and effects data

Higher AVS and TOC concentrations and f_{FP}, frequently correspond with stronger binding and lower toxicity of metals in sediments (Di Toro et al., 2005; Simpson and Batley, 2007). The historical data sets provided wide ranges for these parameters (Table 1). For these sediments the f_{FP} ranged from 2.1% to 100%, TOC from 0.5% to 8.4% and AVS from 0.1 μmol/g to 58 μmol/g. The sediments had HQ values ranging from 0.17 to 9.9 (Table 1), and different metals in the sediments influencing the HQ values (Table S1, Supplementary Material).

The models were first developed using 48 of the 65 historical paired chemistry and ecotoxicity data sets. Owing to the formulation of the models and the resulting mathematical constraints, 17 data sets with 0% and ≥100% reproduction were excluded (noting >100% is possible when expressed as % of control response). The logistic relationship between reproduction of the amphipod, *M. plumulosa*, and HQ and HQ* are described by Eq. (4) and Eq. (5) and presented in Fig. 1. The derivation of HQ* based on HQ, f_{FP}, TOC and AVS, is described by Eq. (6), with HQ* values ranging from 0.18 to 6.23 (Table 1).

$$\log\frac{100-y}{y} = (0.82 \pm 0.22) \bullet \log HQ - (0.60 \pm 0.10) \quad p = 0.0006\ (N = 48) \tag{4}$$

$$\log\frac{100-y}{y} = (1.59 \pm 0.28) \bullet \log HQ - (0.20 \pm 0.13) \bullet \log AVS - (0.27 \pm 0.42) \bullet \log TOC - (0.95 \pm 0.25) \bullet \log f_{FP} + (1.06 \pm 0.36) \quad p < 0.0001\ (N = 48) \tag{5}$$

$$HQ^* = HQ \bullet \left(\frac{AVS}{AVS_{ref}}\right)^{-0.13} \bullet \left(\frac{TOC}{TOC_{ref}}\right)^{-0.17} \bullet \left(\frac{f_{FP}}{f_{FP\,ref}}\right)^{-0.60} = HQ \bullet \left(\frac{AVS}{0.5}\right)^{-0.13} \bullet \left(\frac{TOC}{1.0}\right)^{-0.17} \bullet \left(\frac{f_{FP}}{50}\right)^{-0.60} \quad p < 0.0001\ (N = 48) \tag{6}$$

The HQ* model provided significantly better prediction of toxicity than the HQ model (Fig. 1A and B), as evidenced by an increase in the RMSE (root-mean-square error) of the regression, from 21 (for the HQ model) to 16 (for the HQ* model). A comparison of the measured and predicted toxicity (Figure S2, Supplementary Material) also indicates the improvement, as the R^2 of the measurements relative to the 1:1 line increased from 0.21 to 0.58 when sediment properties were incorporated into the hazard quotient model. The improved HQ* model (Eq. (6)) indicates a greater importance of particle size in influencing toxicity predictions, compared to TOC and AVS. For example, increasing f_{FP} from 10 to 90% results in HQ* increasing by a factor of 3.74, compared to a factor of 1.66 when TOC increases from 0.5 to 10%, and a factor of 1.70 when AVS increases from 0.5 to 30 μmol/g (Table S3). For these historical data there was insufficient measurement of dissolved metals in overlying waters or porewaters to enable TU to be calculated.

3.2. Metal contaminants and release in toxicity tests

The properties of the 26 new test sediments varied considerably (Table 2), with f_{FP} ranging from 32% to 91%, TOC from 1.0% to 18% and AVS from 0.5 μmol/g to 56 μmol/g. The sediment had metal concentrations that ranged from near-background level for the region to highly contaminated, with maximum concentrations of the metals in the

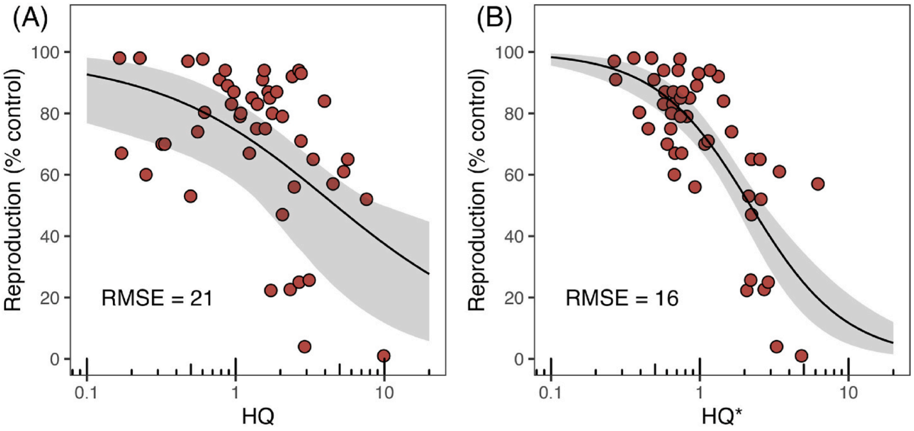


Fig. 1. The logistic relationship between amphipod reproduction and two models: (A) HQ (Eq. (4)) and (B) HQ* (Eq. (5)) for the historical static-renewal sediment toxicity data set (N = 48). The solid curve is the regression line of the measurement using the logistic model (Eq. (1)), and the shaded area represents the 95% confidence band. RMSE is the root-mean-square error of the regression.

sediments being very different (Table S4). The dominant contaminants were Cu, Pb and Zn, with maximum concentrations reaching 1050, 692 and 2680 mg/kg, respectively. One sample had a very high cadmium concentration of 35 mg/kg, in contrast to the remaining samples with cadmium concentrations below 3.2 mg/kg. Concentrations of As, Cr and Ni were relatively low, with maximum values of 26, 116 and 39 mg/kg, respectively. The vast majority of sediment samples exhibited concentrations of Cu, Pb, and/or Zn that exceeded their corresponding SQGV, with the exception of the Control sediment (Sediment-1). Notably, 46% of the sediment samples exceeded upper guideline value (SQG-High) for Cu, while 50% and 65% of the sediment samples exceeded the corresponding SQG-High values for Pb and Zn, respectively. For this data set the HQ values ranged from 0.2 to 6.7, with the highest HQ occurring for sediments with known anthropogenic sources of heavy metals (e.g., Sediment-25 (HQ 6.5) from Blackwattle Bay located within a historically highly-industrialised zone of Sydney Harbour (Cu, Pb, Zn), and Sediment-23 (HQ 6.7) from Cockle Bay, Lake Macquarie, that had 35 mg/kg Cd and high Pb and Zn originating from a smelter) (Table 2). The lowest HQ value was Sediment-1 (HQ = 0.2) from Grays Point near the Royal National Park, which was used as the Control sediment in the toxicity tests.

The dissolved concentrations of Cu and Zn in the overlying water during the 10-d static-renewal and flow-through tests frequently exceeded the WQGVs of 1.3 and 15 μg/L, respectively (Tables S5 and S6, Supplementary Material). In contrast, the dissolved concentrations of As, Cd, Cr, Ni, and Pb were below the limit of detection and also below their respective WQGVs (12 μg As/L from (Golding et al., 2022); 5.5 μg Cd/L, 27.4 μg Cr/L, 70 μg Ni/L, 4.4 μg Pb/L from (ANZG, 2018)), except for Cd in Sediment-9 (10.7 μg/L). The observation that Cu and Zn were the principal metal contaminants in the overlying water was consistent with their high concentrations in the sediments. However, the observation that Pb concentrations remained low in the overlying water, despite its elevated sediment concentrations, can be attributed to the limited release of Pb from sediments (Amato et al., 2016; Remaili et al., 2017). The dissolved metal concentrations in static-renewal tests were significantly higher than in flow-through tests for both Cu ($p < 0.01$) and Zn ($p < 0.01$). The range of TU is 0.3–3.9 for static-renewal tests, and 0.2–1.4 for flow-through tests. The percentage of TU exceeding 1.0 was ~50% for static-renewal tests, while only 8% exceeded 1.0 for flow-through tests. It clearly indicates that the high exchange rate of overlying water in the flow-through tests considerably decreased the concentrations of dissolved metals and the potential toxicity associated with dissolved metal concentrations. In the only published study of dissolved trace metals in Sydney Harbour, Hatje et al. (2003) reported mean dissolved Cd, Cu, Ni, and Zn concentrations of 0.04, 1.7, 0.86, and 6.5 μg/L, respectively, being a similar magnitude for Cu and Zn, but lower for Cd and Ni, compared to what was observed for the flow-through tests (Table S6).

Sulfide in sediments (as represented by AVS) binds Cd, Cu, Ni, Pb and Zn strongly, and thus when AVS concentrations exceed the corresponding molar concentration of these metals, their porewater concentrations are predicted to be negligible (Di Toro et al., 2005), and a low risk of adverse effects to benthic organisms is assessed (USEPA, 2005). The oxidation of surface sediments through abiotic or biotic processes (e.g., bioturbation) may result in oxidation of AVS and increase the efflux of dissolved metals from sediments to overlying waters (Xie et al., 2018) and substantially increase the level of toxicity (Simpson et al., 2012). Among the seven sediments that underwent a resuspension-oxidation treatment, four initially exhibited significant AVS concentrations (Table 2). The reduction in AVS varied across different degrees: marginal (from 12 to 11 μmol/g), moderate (from 18 to 6.9 μmol/g), and substantial (56–13 μmol/g, and 9.2 to 0.5 μmol/g). Interestingly, although the TU was greater for the oxidise sediments, the increase in TU did not correlate with the degree of AVS oxidation (Table 2). This observation underscores the intricate chemical diversity inherent to sediments, where interactions between metals and sediment introduces complexity to the release of metals.

3.3. *Sediment toxicity to M. plumulosa*

Both static-renewal and flow-through toxicity tests were undertaken on the 26 new test sediments (Table 2, Table S7, Supplementary Material). In the static-renewal tests, survival percentages (relative to control) ranged from 58 ± 11% to 105 ± 5%, whereas in the flow-through tests, survival percentages ranged from 84 ± 8% to 107 ± 6% (Figure S3, Table S6). For both tests, reproduction is a more sensitive endpoint than survival. For static-renewal tests, reproduction percentages (relative to control) ranged from 10 ± 4% to 101 ± 4%, while in flow-through tests, they spanned from 33 ± 10% to 114 ± 15%.

There was no significant difference in survival of adult amphipods between the two test types ($p = 0.20$). However, flow-through tests exhibited significantly higher reproduction compared to static-renewal tests ($p < 0.01$). The relative standard deviations of survival and reproduction in flow-through tests ($n = 4$ replicates) were relatively higher than that in the static-renewal tests ($p = 0.03$).

This discrepancy in reproduction between flow-through and static renewal tests was primarily attributed to the considerably lower levels of dissolved metal exposure in flow-through tests compared to static-renewal tests ($p < 0.01$ for TU, Table 2).

For the new static-renewal test data sets ($N = 26$), the fitting was poor for the HQ model (RMSE = 18, Fig. 2A) and also for the HQ* model (RMSE = 19, Fig. 2B), but was markedly improved using the TU model (RMSE = 13, Fig. 2C, Eq. (7)). A comparison of the measured and predicted toxicity based on TU is provided in Figure S4, Supplementary Material.

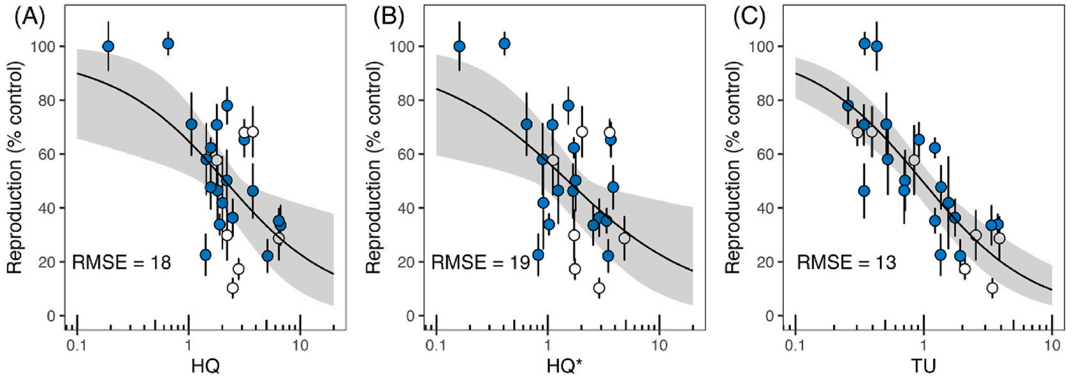


Fig. 2. The logistic relationship between reproduction of amphipod and the three models: (A) HQ (Eq. (4)), (B) HQ* (Eq. (5)), and (C) TU (Eq. (7)) for the new static-renewal sediment toxicity data set ($N = 26$). The open symbols indicate the 7 oxidised sediments (Table 2). The regression fit is described in Fig. 1.

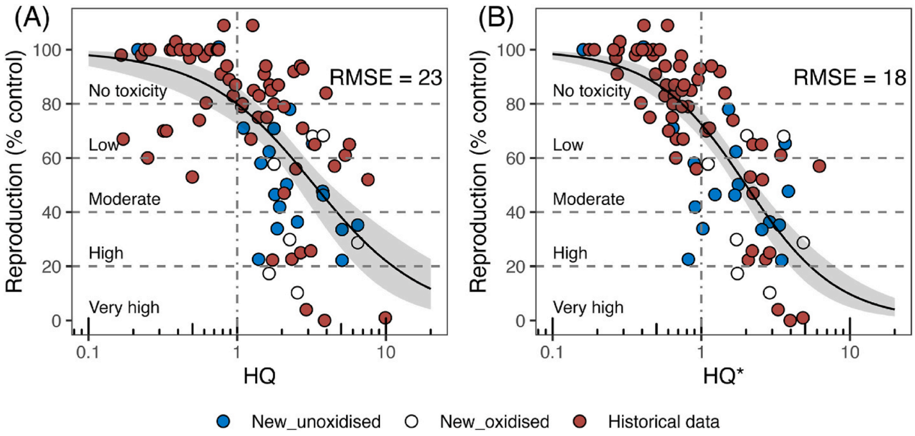


Fig. 3. The logistic relationship between reproduction of amphipod and the models: (A) HQ (Eq. (4)) and (B) HQ* (Eq. (5)) using the combined static-renewal toxicity data sets (N = 74). The regression fit is described in Fig. 1.

$$\log\frac{100-y}{y} = (1.22 \pm 0.20) \bullet \log TU - (0.04 \pm 0.01) \quad p < 0.0001 \ (N = 26) \tag{7}$$

The reduced predictive accuracy of toxicity using the HQ*-based model, developed from the historical data set (N = 48), could potentially be attributed to the narrower HQ* range observed within the smaller data set (0.16–4.85, N = 26), as compared to the historical data set (0.18–6.23, N = 65). Additionally, the inclusion of oxidised sediment treatments (N = 7) might have further contributed to this discrepancy. Analysing the static-renewal test outcomes from both the historical and new data sets in combination (N = 74 paired chemistry and toxicity combinations, Fig. 3), stronger relationships were evident for HQ* (RMSE = 18) in contrast to HQ (RMSE = 23). The comparison between the observed and predicted toxicity, encompassing the combined data, is provided in Figure S5. The combined results indicate that 95% confidence bands for toxicity risk-level predictions by HQ* model will frequently be within 20–30% of the measured reproduction.

3.4. Categories of toxicity

The HQ* model provided significantly better prediction of toxicity than the HQ model, and indicated a greater importance of particle size in influencing toxicity predictions, compared to TOC and AVS. However, it is also useful to consider this improvement in relation to the data use for risk assessment purposes. To evaluate multiple LoE, including chemistry and toxicity data, a weight of evidence (WoE) approach is recommended, and this typically utilises semi-quantitative ranking for each LoE using a numerical-scoring categories (Bay and Weisberg, 2012; Simpson and Batley, 2016).

For the standardised amphipod survival and reproduction test procedure, a system of five toxicity categories was proposed (Table S2). The analysis of the combined 91 measurements indicated that the observed toxicity distribution comprised 45% No Toxicity, 23% Low, 13% Moderate, 13% High, 5% Very High (Table S8, Supplementary Material).

Comparing the predicted toxicity categories, the HQ* model exhibited a 56% accuracy in ‘predicting the measured category of toxicity’, in contrast to 45% achieved by the HQ model (Fig. 4, Table S8). Considering predictions that deviated by one toxicity category from the measured category, the HQ* model demonstrated a 12% higher identification of toxicity and 24% lower identification of toxicity, whereas the HQ model showed a 21% overestimation of toxicity and a 19% underestimation. In sum, the HQ* model achieved 92% accuracy within one category of the measured outcome, as opposed to the 85% accuracy achieved by the HQ model.

3.5. Flow-through tests comparisons

The testing of the same sediments using both static-renewal and flow-through test procedures provided 52 results for which the sediment metal concentrations and bioavailability-influencing properties (f_{FP}, TOC and AVS) were the same in 26 pairs, but the exposure conditions and dissolved metal exposures differed between pairs (static-renewal vs flow-through). The flow-through tests were intended to better resemble

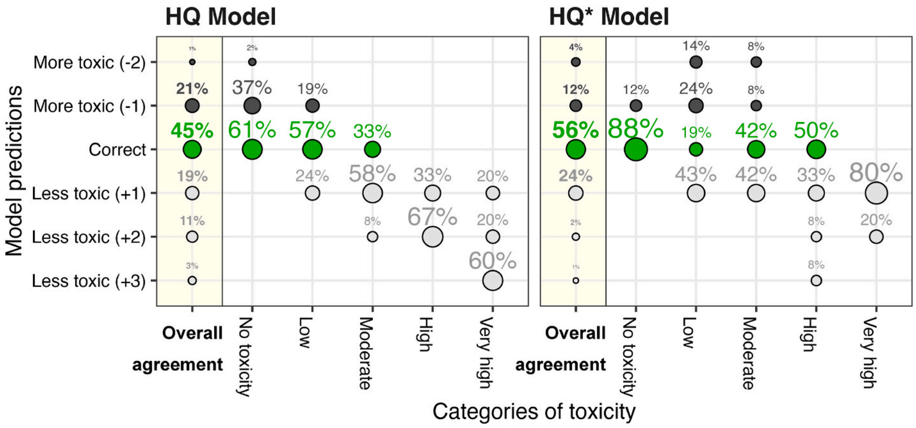


Fig. 4. Agreement between model predictions and measured categories of toxicity: Left series = Overall agreement, and Right series = breakdown for individual categories.

a field environment where the contaminant release to the overlying water is diluted and results in the lower dissolved exposure to the benthic organisms.

The flow-through test results (Figure S6) had poorer fits to all three models (Figure S6, RMSE = 19, 20, and 18 for the HQ, HQ*, and TU models, respectively) when compared to the static-renewal test results (Fig. 2, RMSE = 18, 19, and 13 for the HQ, HQ*, and TU models, respectively). The comparison of the measured toxicity with the TU-model predictions for the static-renewal and flow-through tests highlights the significant contribution of the overlying water metal exposure to the toxicity. The poor fit of the TU-based model for the flow-through data (RMSE = 18, Figure S6C, Supplementary Material) was not unexpected when we consider that the amphipod is a deposit feeder and the major portion of its metal exposure occurs via diet (including sediment ingestion) when the dissolved exposure is low (Campana et al., 2012). The considerably better prediction of toxicity using the TU-based model in the static-renewal tests (RMSE = 13, Fig. 2C) indicates that the detection of highly elevated dissolved metal concentrations makes risk prediction easier, but the translation of the toxicity risk predictions from in laboratory and field exposure scenarios in less certain.

3.6. Implications for toxicity data as a line of evidence for risk assessment

The collection and evaluation of multiple LoE is expected to improve the quality of the decision making (Bay and Weisberg, 2012; Simpson and Batley, 2016). However, the cost of gathering the LoE must be balanced against the values gained for the decision process. Our reasoning for this study, was a desire to be able to better predict risks of adverse effect of contaminants using routine and cost-effective measurements. The primary aim of this study was to investigate whether the prediction of toxicity risks based on the concentrations of major contaminant expressed in the form of a hazard quotient (HQ) could be significantly improved through the consideration of common sediment properties that modify contaminant bioavailability (particle size, TOC, AVS). When considered as categories of risk of adverse effects (from low to high), the Chemistry-LoE (HQ-based) and Ecotoxicology-LoE had better agreement when contaminant bioavailability-modifying factors were included for the Chemistry-LoE. Overall, we found that modest improvements of toxicity risk prediction could be achieved and that the most influential factor was sediment particle size.

The Ecotoxicology-LoE relies on the use of surrogate organisms, and we assessed toxicity to the reproduction of an epibenthic deposit feeding amphipod, *M. plumulosa* (Mann et al., 2009). This represents a single surrogate organism upon which risk assessment decisions may be influenced, and other organisms will have differing sensitivity, update and exposure routes for contaminants (Rainbow, 2007). The value of the Ecotoxicology-LoE come with the observation of toxicity, as if toxic effects are observed for the tested species if is likely that many other species of organism will exist in the environment that may be impacted by the same contaminant exposure. The observation of no toxicity is also useful, but questions remain about risks of adverse effect to organisms that quite different to the surrogate species used for the Ecotoxi cology-LoE.

The comparison of the static-renewal and flow-through test results raise important questions concerning the more common use of static or static-renewal toxicity test results for risk assessment decisions. Laboratory-based tests may frequently not adequately resemble conditions of organism exposure at the field location (Burton et al., 2005; Harrison et al., 2019; Mann et al., 2010), and this can lower the value of the ecotoxicology data as a line of evidence. The design of both test procedures can be questioned relative to field-exposure conditions. It is possible that the taller flow-through test chambers create a sheltered environment that allows the amphipods, that are epibenthic, to potentially avoid the porewater and dissolved metal flux exposure more than they could in the field by clinging on to the walls of the chamber above the sediment. Providing the ability for organisms to avoid sediment contaminant exposure may be very relevant for some organisms (Araújo et al., 2016; Ward et al., 2013), but it may also be artificial for others (e.g., where not burrowing would result in exposure to predators). The design of the static-renewal tests that keeps the amphipods closer to the sediments and creating a higher dissolved metal exposure provides a more conservative LoE for toxicity.

4. Conclusions

This study reaffirms that consideration of factors that influence contaminant bioavailability improves toxicity risk predictions, however the improvements may be modest. Only the first study hypothesis was proven, where the HQ modified using sediment particle size, TOC and AVS improved toxicity risk predictions for static renewal tests. Notably, sediment particle size emerges as the most pivotal factor in this refinement. However, this enhancement is not mirrored in flow-through data sets, and outcomes for sediments with AVS reduced through oxidation remain inconclusive.

While dissolved metals in overlying waters prominently influence toxicity outcomes in static-renewal tests, their impact diminishes in flow-through tests. This disparity indicates that toxicity risk predictions derived from static and static-renewal tests may overstate the actual risk of sediment toxicity in field settings characterized by high water-exchange rates. Consequently, the translation of the toxicity risk projections from laboratory settings to field conditions demands meticulous scrutiny, warranting careful assessment before informing regulatory determinations.

Author statement

Yanfeng Zhang: Conceptualization, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. Minwei Xie: Formal analysis, Writing – original draft, Writing – review & editing. David M. Spadaro: Investigation; Stuart L. Simpson: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank Josh King and Chad Jarolimek for assistance with instrumental methods of analysis. This study was supported by the National Natural Science Foundation of China (Grant No.42077372).

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2023.122708.

References

Amato, E.D., Simpson, S.L., Jarolimek, C.V., Jolley, D.F., 2014. Diffusive gradients in thin films technique provide robust prediction of metal bioavailability and toxicity in estuarine sediments. Environ. Sci. Technol. 48, 4485–4494.

Amato, E.D., Simpson, S.L., Remaili, T.M., Spadaro, D.A., Jarolimek, C.V., Jolley, D.F., 2016. Assessing the effects of bioturbation on metal bioavailability in contaminated sediments by diffusive gradients in thin films (DGT). Environ. Sci. Technol. 50, 3055–3064.

ANZG, 2018. Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Governments and Australian State and Territory Governments. Canberra ACT, Australia.

Araújo, C.V., Moreira-Santos, M., Ribeiro, R., 2016. Active and passive spatial avoidance by aquatic organisms from environmental stressors: a complementary perspective and a critical review. Environ. Int. 92, 405–415.

Bay, S.M., Weisberg, S.B., 2012. Framework for interpreting sediment quality triad data. Integrated Environ. Assess. Manag. 8, 589–596.

Birch, G., 2017. Assessment of human-induced change and biological risk posed by contaminants in estuarine/harbour sediments: Sydney Harbour/estuary (Australia). Mar. Pollut. Bull. 116, 234–248.

Burton, G.A., Greenberg, M.S., Rowland, C.D., Irvine, C.A., Lavoie, D.R., Brooker, J.A., Moore, L., Raymer, D.F.N., McWilliam, R.A., 2005. In situ exposures using caged organisms: a multi-compartment approach to detect aquatic toxicity and bioaccumulation. Environ. Pollut. 134, 133–144.

Campana, O., Blasco, J.N., Simpson, S.L., 2013. Demonstrating the appropriateness of developing sediment quality guidelines based on sediment geochemical properties. Environ. Sci. Technol. 47, 7483–7489.

Campana, O., Simpson, S.L., Spadaro, D.A., Blasco, J., 2012. Sub-lethal effects of copper to benthic invertebrates explained by sediment properties and dietary exposure. Environ. Sci. Technol. 46, 6835–6842.

Chariton, A.A., Roach, A.C., Simpson, S.L., Batley, G.E., 2010. Influence of the choice of physical and chemistry variables on interpreting patterns of sediment contaminants and their relationships with estuarine macrobenthic communities. Mar. Freshw. Res. 61, 1109–1122.

Ciutat, A., Boudou, A., 2003. Bioturbation effects on cadmium and zinc transfers from a contaminated sediment and on metal bioavailability to benthic bivalves. Environ. Toxicol. Chem.: Int. J. 22, 1574–1581.

Dafforn, K.A., Simpson, S.L., Kelaher, B.P., Clark, G.F., Komyakova, V., Wong, C.K.C., Johnston, E.L., 2012. The challenge of choosing environmental indicators of anthropogenic impacts in estuaries. Environ. Pollut. 163, 207–217.

Di Toro, D.M., McGrath, J.A., Hansen, D.J., Berry, W.J., Paquin, P.R., Mathew, R., Wu, K. B., Santore, R.C., 2005. Predicting sediment metal toxicity using a sediment biotic ligand model: methodology and initial application. Environ. Toxicol. Chem. 24, 2410–2427.

Golding, L.A., Valdivia, M.V., van Dam, J.W., Batley, G.E., Apte, S.C., 2022. Toxicity of arsenic (v) to temperate and tropical marine biota and the derivation of chronic marine water quality guideline values. Environ. Chem. 19, 116–131.

Harrison, A.M., Hudson, M.L., Burton Jr., G.A., 2019. Hyporheic interactions increase zinc exposure and effects on hyalella azteca in sediments under flow-through conditions. Environ. Toxicol. Chem. 38, 2447–2458.

Hatje, V., Apte, S.C., Hales, L.T., Birch, G.F., 2003. Dissolved trace metal distributions in port jackson estuary (sydney harbour), Australia. Mar. Pollut. Bull. 46 (6), 719–730.

Long, E.R., 2006. Calculation and uses of mean sediment quality guideline quotients: a critical review. Environ. Sci. Technol. 40, 1726–1736.

Mann, R.M., Hyne, R.V., Simandjuntak, D.L., Simpson, S.L., 2010. A rapid amphipod reproduction test for sediment quality assessment: in situ bioassays do not replicate laboratory bioassays. Environ. Toxicol. Chem. 29, 2566–2574.

Mann, R.M., Hyne, R.V., Spadaro, D.A., Simpson, S.L., 2009. Development and application of a rapid amphipod reproduction test for sediment-quality assessment. Environ. Toxicol. Chem.: Int. J. 28, 1244–1254.

Maruya, K.A., Landrum, P.F., Burgess, R.M., Shine, J.P., 2012. Incorporating contaminant bioavailability into sediment quality assessment frameworks. Integrated Environ. Assess. Manag. 8, 659–673.

Rainbow, P.S., 2007. Trace metal bioaccumulation: models, metabolic availability and toxicity. Environ. Int. 33, 576–582.

Remaili, T.M., Simpson, S.L., Jolley, D.F., 2017. Effects of enhanced bioturbation intensities on the toxicity assessment of legacy-contaminated sediments. Environ. Pollut. 226, 335–345.

Schlekat, C.E., Garman, E.R., Vangheluwe, M.L., Burton Jr., G.A., 2016. Development of a bioavailability-based risk assessment approach for nickel in freshwater sediments. Integrated Environ. Assess. Manag. 12, 735–746.

Simpson, S., Batley, G., 2016. Sediment Quality Assessment: a Practical Guide. CSIRO Publishing, Melbourne, Victoria, Australia.

Simpson, S.L., Batley, G.E., 2007. Predicting metal toxicity in sediments: a critique of current approaches. Integrated Environ. Assess. Manag. 3, 18–31.

Simpson, S.L., Spadaro, D.A., Batley, G.E., Irvine, I.A., Synnot, R.N., 2020. Remediation criteria for gasworks-impacted sediments: assessing the effects of legacy hydrocarbons and more recent metal contamination. Sci. Total Environ. 737, 139725.

Simpson, S.L., Spadaro, D.A., O'Brien, D., 2013. Incorporating bioavailability into management limits for copper in sediments contaminated by antifouling paint used in aquaculture. Chemosphere 93, 2499–2506.

Simpson, S.L., Spadaro, D.A., Watters, A., 2014. Newcastle Port Corporation Port-wide Strategy – Future Capital Dredging Assessment. CSIRO Wealth from Oceans Report EP143733, p. 549.

Simpson, S.L., Ward, D., Strom, D., Jolley, D.F., 2012. Oxidation of acid-volatile sulfide in surface sediments increases the release and toxicity of copper to the benthic amphipod Melita plumulosa. Chemosphere 88, 953–961.

Spadaro, D.A., Simpson, S.L., 2016. Appendix F: Protocol for Whole-Sediment Sub-lethal (Reproduction) Toxicity Tests Using the Copepod Nitocra Spinipes (Harpacticoid). SEDIMENT QUALITY ASSESSMENT, p. 276.

USEPA, 2005. Procedures for the Derivation of Equilibrium Partitioning Sediment Benchmarks (ESBs) for the Protection of Benthic Organisms: Metal Mixtures (Cadmium, Copper, Lead, Nickel, Silver, and Zinc). US Environmental Protection Agency, Office of Research and Development, Washington, DC, USA.

Ward, D.J., Simpson, S.L., Jolley, D.F., 2013. Avoidance of contaminated sediments by an amphipod (Melita plumulosa), a harpacticoid copepod (Nitocra spinipes), and a snail (Phallomedusa solida). Environ. Toxicol. Chem. 32, 644–652.

Xie, M., Simpson, S.L., Huang, J., Teasdale, P.R., Wang, W.-X., 2021. In Situ DGT sensing of bioavailable metal fluxes to improve toxicity predictions for sediments. Environ. Sci. Technol. 55, 7355–7364.

Xie, M., Simpson, S.L., Wang, W.-X., 2019. Bioturbation effects on metal release from contaminated sediments are metal-dependent. Environ. Pollut. 250, 87–96.

Xie, M., Wang, N., Gaillard, J.-F., Packman, A.I., 2018. Interplay between flow and bioturbation enhances metal efflux from low-permeability sediments. J. Hazard Mater. 341, 304–312.

Zhang, Y., Spadaro, D.A., King, J.J., Simpson, S.L., 2020. Improved prediction of sediment toxicity using a combination of sediment and overlying water contaminant exposures. Environ. Pollut. 266, 115187.

Supplementary Material

Improving toxicity prediction of metal-contaminated sediments by incorporating sediment properties

Yanfeng Zhang [a,b], Minwei Xie [c]*, David M. Spadaro [a], Stuart L. Simpson [a,d]

[a] *CSIRO Environment, Lucas Heights, NSW 2234, Australia*

[b] *Tianjin Key Laboratory of Remediation & Pollution Control for Urban Ecological Environment, Key Laboratory of Pollution Processes and Environmental Criteria of Ministry of Education, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China*

[c] *State Key Laboratory of Marine Environmental Science, Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, College of the Environment and Ecology, Xiamen University, Xiamen, Fujian 361102, China*

[d] *Hong Kong Branch of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China*

* To whom correspondence may be addressed (minweixie@xmu.edu.cn)

S1. Methods

S1.1 Historical sediment chemistry and toxicity data sets

Previous studies of sediment quality that included chemistry and ecotoxicology assessment using the amphipod toxicity tests were identified and relevant data extracted. Data from four historical studies were used: Amato et al. (2014), and Simpson et al. (2013, 2014 and 2020), where in each case metals were the dominant sediment contaminants. These studies provided 65 paired chemistry and toxicity data sets (Table 1; Table S1).

Table S1. Historical sediment data overview: total recoverable metal concentrations (mg/kg dry weight). Sediment identifiers align with those in Table 1.

Sediment	As	Cd	Cr	Cu	Ni	Pb	Zn	Sediment	As	Cd	Cr	Cu	Ni	Pb	Zn
1	7.3	0.2	19	3.4	2.2	8.1	23	**34**	21	0.6	47	**790**	23	20	230
2	5.0	0.1	12	7.5	3.0	15	41	**35**	22	0.2	51	**810**	22	21	310
3	8.6	0.2	24	5.8	3.2	14	44	**36**	14	1.3	49	40	31	**248**	**1880**
4	4.0	0.1	9.9	30	1.0	30	50	**37**	5.9	**18**	6.9	37	2.7	177	**431**
5	5.7	0.1	18	8.8	18	7	107	**38**	20	0.7	31	**975**	15	16	271
6	3.9	0.1	17	35	8.9	9.7	145	**39**	20	0.7	31	**979**	15	16	271
7	15	0.2	25	17	7.7	41	144	**40**	13	1.4	47	126	12	**255**	**2120**
8	10	0.2	27	27	11	49	120	**41**	**70**	0.6	55	**835**	33	44	401
9	6.0	0.4	15	110	6.0	21	148	**42**	22	0.8	34	**1200**	16	14	340
10	11	0.7	43	23	20	38	130	**43**	17	2.2	71	53	44	**410**	**2670**
11	9.2	0.5	48	39	23	23	156	**44**	19	0.6	19	**1759**	14	13	405
12	15	0.6	35	143	16	17	115	**45**	19	0.7	25	**2120**	13	13	**415**
13	3.8	0.4	67	71	8.5	99	240	**46**	21	0.7	28	**2270**	13	13	**415**
14	5.9	0.1	6.9	9.6	2.9	163	**437**	**47**	20	0.6	32	**3080**	15	18	**558**
15	4.6	0.3	33	23	31	53	**563**	**48**	**110**	**35**	30	220	15	**850**	**3800**
16	8.2	0.4	77	75	10	110	270	***49***	4.6	0.1	20	8.2	2.2	23	83
17	16	1.1	70	53	8.4	115	325	***50***	4.2	0.5	13	9.0	7.0	21	42
18	11	0.8	84	81	10	110	310	***51***	5.0	0.1	23	15	5.5	37	133
19	16	0.6	22	**368**	11	12	160	***52***	4.8	0.1	30	15	5.5	37	133
20	6.7	0.4	95	120	13	150	360	***53***	13	0.2	24	13	6.1	31	109
21	12	0.6	46	31	34	88	**857**	***54***	6.0	0.6	18	64	6.0	5.0	120
22	23	0.8	54	**310**	26	22	200	***55***	12	0.5	19	22	9.0	41	100
23	18	0.6	56	**400**	28	20	150	***56***	9.1	0.5	26	26	13	50	120
24	13	0.4	110	140	16	180	**420**	***57***	19	0.2	29	25	11	60	216
25	11	1.2	59	49	47	143	**626**	***58***	19	0.2	29	25	11	60	216
26	18	0.9	31	**500**	15	14	211	***59***	19	0.5	30	38	13	68	180
27	3.0	0.4	130	170	17	200	**480**	***60***	19	0.8	29	35	12	65	200
28	5.0	3.0	23	55	11	110	**1200**	***61***	6.6	0.6	33	25	28	61	328
29	11	1.4	58	49	48	184	**682**	***62***	10	0.5	59	48	**55**	75	403
30	12	0.8	57	46	**53**	137	**1080**	***63***	2.7	0.4	93	130	12	150	350
31	13	2.2	64	43	45	**220**	**861**	***64***	11	1.1	62	39	49	110	**482**
32	5.8	**14**	6.5	8.5	2.3	127	354	***65***	10	5.8	40	110	17	**260**	**2900**
33	4.4	3.5	25	80	13	180	**1600**								
SQGV	20	1.5	80	65	21	50	200	SQGV	20	1.5	80	65	21	50	200
SQG-High	**70**	**10**	**370**	**270**	**52**	**220**	**410**	SQG-High	**70**	**10**	**370**	**270**	**52**	**220**	**410**

SQGV: sediment quality guideline value for protection of 95% marine water species (Simpson et al., 2013b).

SQG-High: high guideline value for sediment quality (Simpson et al., 2013b).

Concentrations are underlined when between SQGV and SQG-High, and **bold** when above SQG-High.

S1.2 New sediment collection and sediment manipulation

The sediments were collected from intertidal and subtidal (<2 m water depth) locations from sites in the Sydney region (New South Wales, Australia) using a shovel or grab sampler to provide approximately 4 kg of surface sediment (0-10 cm depth). Clean sediment was collected from the Royal Nation Park (New South Wales, Australia) and was used as the control in the toxicity tests. The sampling sites were selected based on our previous investigations to guarantee that metals were the principal contaminants. Sediments were sieved through a 2 mm plastic mesh by pressing to remove large debris, and then homogenised and stored in plastic bottles in cold room (4 °C). In order to minimise sample oxidation or other changes, sediment analyses and toxicity tests were generally conducted within 8 weeks (Simpson and Batley, 2016).

Out of the 19 newly collected sediment samples, 7 underwent a partial oxidation process to yield sediments with equivalent metal concentrations but reduced AVS levels. This transformation was achieved by blending 1 kg of sediment with 200 mL of clean seawater within a 2-liter plastic container. The container was gently rotated on a roller mixer, operating at a speed of 2 rounds per minute. Over a span of 21 days, the sediment oxidation process ensued, punctuated by intermittent settling periods (lasting 1 to 2 hours) every two days, during which the overlying seawater was refreshed. The oxidation process was facilitated by the oxygen present in the free headspace of the container and the overlying water. Concurrently, the regular renewal of the water served to avert the acidification of the sediment slurry.

S1.3 Chemical analyses

All chemicals were of AR or GR grade. Ultrapure water (18 MΩ·cm, Milli-Q, Millipore) was used for preparation of reagents. All the devices and vessels used in analyses were washed with acid (5% HNO_3) and rinsed using ultrapure water. For analyses of physicochemical parameters of overlying water, dissolved oxygen, temperature, salinity and pH were measured using a portable multiple-parameter monitoring device (WTW, Wissenschaftlich-Technische Werstätten). Dissolved ammonia was analysed colorimetrically using a Merck Spectroquant Kit (No. 14752).

Methods for analyses of sediment properties and contaminants have been described in detail in our previous works (Simpson et al., 2013a; Simpson and Batley, 2016). Particle size distribution (% <63 µm, recorded as fraction of fine particles, f_{FP}) was determined by Mastersizer 3000 (Malvern Panalytical). Total organic carbon (TOC) was analysed using the method of high temperature combustion coupled infrared detection with TOC-V Total Organic Carbon Analyser

(Shimadzu). Acid volatile sulfide (AVS) was determined using a rapid method that measured the concentration of sulfides released from the sediment using dilute acid (1 mol L^{-1} H^+) (Simpson, 2001).

Total recoverable metals (TRM) were analysed following low pressure aqua regia digestion of sediment in a microwave digestion system (MARS 5, CEM) and dilute-acid-extractable metals (AEM) were obtained by wet sediment digestion in 1 mol/L HCl for 60 min. Overlying water was collected by filtration (0.45 µm) and acidized with concentrated HNO_3 (0.2%, v/v). The analyses of metals in waters and acid digests were performed using inductively coupled plasma – atomic emission spectrometry (Agilent 720 ICP-OES, SPS 3 autosampler). For quality assurance/quality control (QA/QC) purposes, analyses included blanks, duplicates and certified reference materials (CRM) for metals (ERM®-CC018, European Reference Materials, and results were within 30% for duplicates and with 75 – 125% of the certified values for CRMs.

Analyses of polycyclic aromatic hydrocarbons (PAHs) and total recoverable hydrocarbons (TRHs) were undertaken by the National Measurement Institute (NMI, Sydney, Australia), a laboratory accredited by the National Association of Testing Authorities (NATA) of Australia, as described by Simpson et al. (2020). Total PAHs comprise 16 compounds: naphthalene, acenaphthylene, acenaphthene, fluorene, phenanthrene, anthracene, fluoranthene, pyrene, benzo(a)anthracene, chrysene, benzo(b)fluoranthene, benzo(k)fluoranthene, benzo(a)pyrene, indeno(1,2,3-cd)pyrene, dibenzo(a,h)anthracene and benzo(g,h,i)perylene. The PAHs were extracted from sediments using dichloromethane/acetone, and the TRHs were extracted with either methanol followed by spiking into water (volatile C6-C10 TRHs) or dichloromethane/acetone (semi-volatile C11-C36 TRHs and PAHs). Determinations were conducted by capillary gas chromatography–mass spectrometry (GC–MS). For QA/QC purposes, spike-recoveries of surrogates for organics were accepted when within 90-130%. Maximum acceptable relative percent difference on spikes and duplicates was 30% for metals and 40% for organics.

S1.4 Sediment toxicity tests

The tested species, *Melita plumulosa,* is an epibenthic amphipod from estuaries of south-eastern Australia and continuously cultured in our laboratory. Clean seawater used in the tests was collected from the south coast of Sydney of Australia, filtered (1 µm) and diluted to 30‰ salinity with deionized water. The seawater was analysed prior to use to ensure no contaminants were present at significant concentrations.

Toxic effects to the survival and reproduction of *M. plumulosa* were assessed over a period of 10 days using two procedures that created different exposure environments, a standard static-renewal test procedure (Simpson and Batley, 2016) and a flow-through test procedure that enabled much greater flow rates of the overlying water (Zhang et al., 2020; Xie et al., 2021).

In the static-renewal tests, the organisms were exposed to contaminants in 250 mL glass beakers that contained 40 g of sediment and 200 mL of overlying seawater. Four replicates were set for every test. Samples of overlying water were collected on the first and the last day of the tests, as well as on Days 3, Day 5 and Day 7 when the overlying water was exchanged with clean seawater. Time-weighted average concentrations were obtained from the average of the five measurements. Sediment renewal was carried out on Day 5. The standard tests were conducted at constant temperature (21 ± 1 °C) in an environmental chamber (Labec Refrigerated Cycling Incubator, Laboratory Equipment) on a 12-h light/12-h dark rotational cycle and aeration was provided.

The flow-through tests were conducted in exposure chambers that allowed for seawater exchange within the surroundings and made dilution of contaminants released from sediment into overlying water (Figure S1). The exposure chamber was constructed from Perspex cylinders (550 mL, 5.5 cm internal diameter (ID) and 23 cm height) and contained about 4 cm depth of sediment by a cap at the bottom. The test chamber had two ports (2 mm ID) on the side for addition of amphipod food and water entry, and a mesh (180 µm Nylon) at the top for water to exit the chamber. Four replicates of chambers with the same sediment were submerged within a 110-L plastic tank (52 cm length × 42 cm width × 52 cm height) that contained 75 L of seawater. A pump circulated the seawater from the top of the tank and pumped it back into the exposure chambers through the port. The rate of seawater flow into each chamber was approximately 240 ± 30 mL/min, resulting in displacement of the chamber volume approximately every 2 to 2.5 minutes and 600-700 times per day. Samples of overlying water were collected on the first and the last day of the tests, as well as on Days 3 and Day 7. Time-weighted average concentrations were calculated from the average of the four measurements. The flow-through tests were conducted in an air-conditioned laboratory (21 ± 1 °C) with normal day-light cycles and aeration provided with the water recycling.

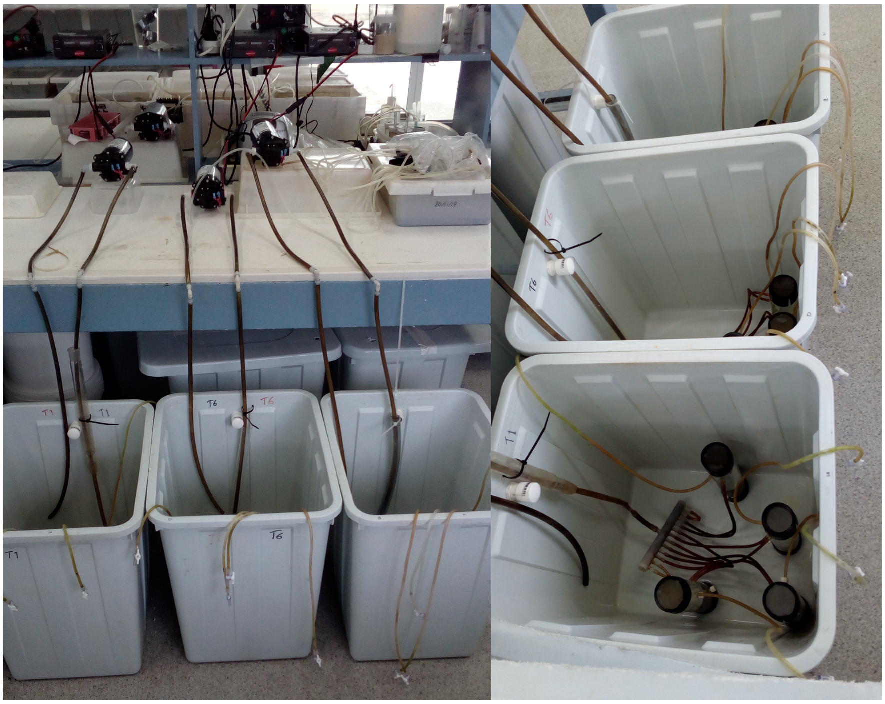


Figure S1. The equipment for the flow-through sediment tests.

The devices such as beakers and chambers were washed before use in a dishwasher (Gallay Scientific) followed by rinsing with reverse osmosis-purified water. Sediments were re-homogenised immediately prior to being added to the test containers, and filtered seawater was added with care to minimise sediment resuspension. Sediments were equilibrated for 48 h before the test started. At the start of the tests, the overlying seawater was renewed and six gravid females (gravid for <24 h) and six males (isolated from laboratory cultures) were randomly assigned to each test chamber. All treatments (static-renewal and flow-through) were fed at a rate of 0.5 mg Sera Micron (Heinsberg, Germany) fish food per amphipod twice a week (on Day 1 and Day 5). To ensure that physicochemical parameters remained within acceptable ranges, measurements of the overlying water were undertaken at the beginning and periodically throughout the tests, including dissolved oxygen (>85% saturation), pH (8.0 ± 0.2), salinity (30 ± 2‰), temperature (21 ± 1°C) and ammonia (<3 mg NH_3−N/L).

The percentage of survival and reproduction of *M. plumulosa* compared with the control were recorded in the two methods. In the static-renewal method, the adults were gently sieved away on Day 5 and placed into the renewed sediment that had been equilibrated for 48 h. This allowed for the removal of juveniles from the first brood which was typically unaffected by contaminants in the test sediment. There was no sediment renewal in the flow-through method. For both tests, on Day 10 the survival of adults was recorded, and the number of embryos was counted under microscopy. Juvenile amphipods were also collected by sieving the sediment through 180 µm mesh after fixing with formaldehyde and staining with Rose Bengal (Spadaro and Simpson, 2016). The total number of embryos and juveniles were summed and the number of offspring per female was recorded. For quality assurance purposes, 8 to 16 juveniles per female were required in all controls for tests to be considered acceptable.

A sediment is considered to be acutely toxic if the survival as a percentage of the control is <80% and is statistically significantly less ($P<0.05$) than the controls. Chronic toxicity is detected when the reproductive output percent control is <85% and is statistically significant less ($P<0.05$) than the controls. The categories of acute and chronic toxicity to the amphipods are defined the Table S2.

Table S2 Categories of toxicity for the amphipod (*Melita plumulosa)*, survival and reproduction test

Category	Survival (Acute toxicity)	Reproduction (Chronic toxicity)
No toxicity	>80% survival at day 10 termination	>80% reproduction at day 10 termination
Low	>65-80% survival at day 10 termination	>60-80% reproduction at day 10 termination
Moderate	>50-65% survival at day 10 termination	>40-60% reproduction at day 10 termination
High	>25%-50% survival at day 10 termination	>20-40% reproduction at day 10 termination
Very high	0-25% survival at Day 5 renewal	0-20% reproduction at day 10 termination

Categories based on results of n = 4 replicated toxicity test expressed as %-control. The lower of the Acute and Chronic categories used as overall category, and preferably the sub-lethal Reproduction category

S1.5 Data analyses

Hazard quotients (HQ) for the sediments were calculated based on the contaminants concentrations of the sediment as described previously (Long et al., 2006; Simpson et al., 2013a) using each sediment quality guideline value (SQGV) (Table S1) (Simpson et al., 2013b). As the 26 sediments were mainly metals-contaminated, HQ was calculated using concentrations of the seven metals (M) of As, Cd, Cu, Cr, Ni, Pb, and Zn: $HQ = \sum_n([M]/SQGV)/n$, where [M] is the metal concentration in sediment, SQGV is the corresponding sediment quality guideline value, and n is the number of contaminants (i.e. 7 metals in this case).

Toxic units (TU) for overlying waters were calculated as the sum of time-weighted average concentrations of dissolved metals (dM) in the overlying waters divided by their respective marine water quality guideline value (Table S2) (WQGV) (ANZG, 2018), i.e. $TU = \sum_n(dM/WQGV)/n$, where dM is the metal concentration in overlying water, WQGV is the corresponding water quality guideline value. Since the seawater WQGV for As is not available, the number of metals here is six (i.e. n = 6). Values of HQ and TU <1 indicate a low risk of toxicity, and values >1 indicate increasing risk and potential magnitude of toxicity as the values increase.

The relationship between the toxicity to amphipod reproduction and the derived models (HQ, HQ*, or TU) were assessed by regression using the logistic model (Equation 1, main text). This was performed in R (version 4.2.2) using the nls() function. The goodness of the fit was assessed by a root-mean-square error (RMSE) parameter. The produced curve and shaded area (Figures 1, 2 3 and S6) represent the regression line and the 95% confidence band, which were generated by the predictNLS() function of the "propagate" package in R.

When assessing the 'goodness of fit' (prediction) using different models (HQ, HQ*, and TU), the measured results and predicted results were compared relative to the 1:1 line (as shown in Figures S2, S4, and S5). An R^2 parameter was calculated using the results relative to the 1:1 line using the following equation:

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = 1 - \frac{\sum_i (x_i - y_i)^2}{\sum_i (\bar{y} - y_i)^2}$$

where x_i and y_i are the x- and y-coordinate of a point in the coordinate system in Figure S2, S4, and S5, $\bar{y}$ is the mean of y_i.

References

ANZG, 2018. Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Australian and New Zealand Governments and Australian state and territory governments, Canberra ACT, Australia. www.waterquality.gov.au/anz-guidelines.

Golding, L.A., Valdivia, M.V., van Dam, J.W., Batley, G.E., Apte, S.C., 2022. Toxicity of arsenic(V) to temperate and tropical marine biota and the derivation of chronic marine water quality guideline values. Environ. Chem. 19, 116-131.

Long, E.R., Ingersoll, C.G. and Macdonald, D.D., 2006. Calculation and uses of mean sediment quality guideline quotients: A critical review. Environ. Sci. Technol. 40, 1726–1736.

Simpson, S.L., 2001. A rapid screening method for acid volatile sulfide in sediments. Environ. Toxicol. Chem. 20, 2657–2661.

Simpson, S.L., Spadaro, D.A., O'Brien, D., 2013a. Incorporating bioavailability into management limits for copper and zinc in sediments contaminated by antifouling paint and aquaculture. Chemosphere, 93, 2499–2506.

Simpson, S.L., Batley, G.E., Chariton, A.A., 2013b. Revision of the ANZECC/ARMCANZ sediment quality guidelines. CSIRO Land and Water Science Report 08/07. Lucas Heights, NSW Australia.

Simpson, S.L. and Batley, G.E., 2016. Sediment Quality Assessment: A Practical Guide. CSIRO Publishing, Melbourne, Victoria.

Simpson, S.L., Spadaro, D.A., Batley, G.E., Irvine I.A., Synnot R.N., 2020. Remediation criteria for gasworks-impacted sediments: Assessing the effects of legacy hydrocarbons and more recent metal contamination. Sci. Total Environ. 737, 139725.

Xie, M., Simpson, S.L., Huang, J., Teasdale P.R., Wang, W.X., 2021. In situ DGT sensing of bioavailable metal fluxes to improve toxicity predictions for sediments. Environ. Sci. Technol. 55, 7355-7364.

Zhang, Y., Spadaro, D.A., King, J.J., Simpson, S.L., 2020. Improved prediction of sediment toxicity using a combination of sediment and overlying water contaminant exposures. Environ. Pollut. 266, 115187.

S2. Results and analysis

Toxicity models derived using historical chemistry and effects data

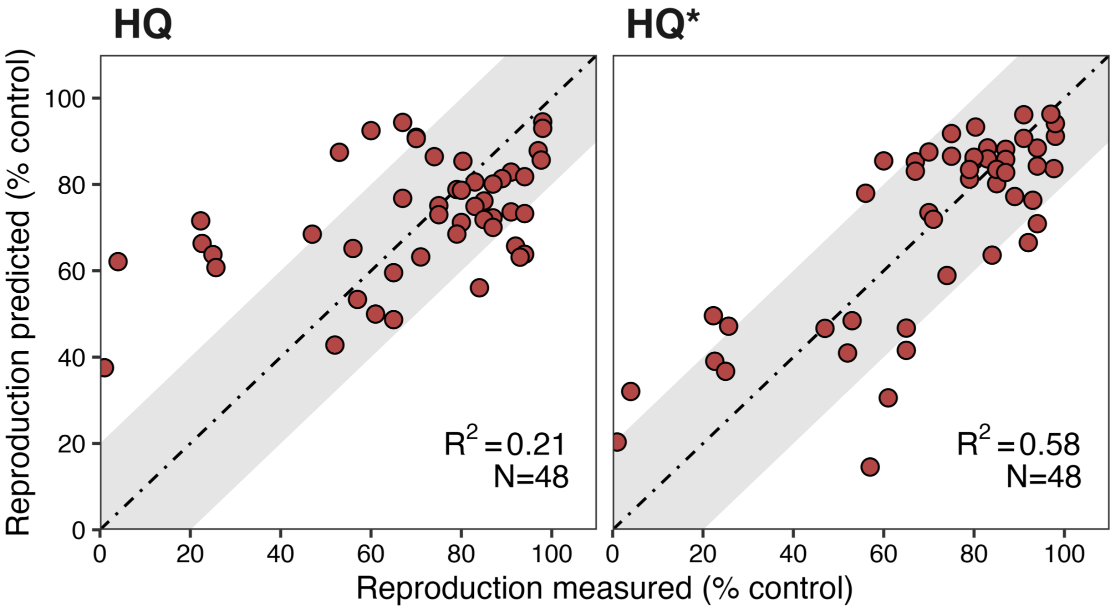


Figure S2. Measured and predicted amphipod reproduction comparison for the two models: HQ (Eq. 4) and HQ* (Eq. 5) for the historical static-renewal sediment toxicity data set. The dot-dashed line represents the 1:1 line, and the shaded band represents a factor of two.

Table S3 Equation 6 that describes relationship between HQ and HQ*, as modified by AVS, TOC and f_{FP}, is used to illustrate the influence of these parameters on the resulting HQ*. HQ* = HQ = 1 when AVS=0.5 µmol/g, TOC=1% and f_{FP}=50%.

$$HQ^* = HQ \cdot \left(\frac{AVS}{0.5}\right)^{-0.13} \cdot \left(\frac{TOC}{1.0}\right)^{-0.17} \cdot \left(\frac{f_{FP}}{50}\right)^{-0.60}$$ (Equation 6, main text)

Below: AVS constant (0.5) and f_{FP} and TOC variable				
AVS (µmol/g) = 0.5	**TOC (%)**			
	0.5	**1**	**3**	**10**
f_{FP} (%)	HQ* value derived when HQ = 1			
10	2.95	2.63	2.18	1.78
50	1.13	**1.00**	0.83	0.68
70	0.92	0.82	0.68	0.55
90	0.79	0.70	0.58	0.48
Below: AVS constant (3) and f_{FP} and TOC variable				
AVS (µmol/g) = 3	**TOC (%)**			
	0.5	**1**	**3**	**10**
f_{FP} (%)	HQ* value derived when HQ = 1			
10	2.34	2.08	1.73	1.41
50	0.89	0.79	0.66	0.54
70	0.73	0.65	0.54	0.44
90	0.63	0.56	0.46	0.38
Below: AVS constant (10) and f_{FP} and TOC variable				
AVS (µmol/g) = 10	**TOC (%)**			
	0.5	**1**	**3**	**10**
f_{FP} (%)	HQ* value derived when HQ = 1			
10	2.00	1.78	1.48	1.20
50	0.76	0.68	0.56	0.46
70	0.62	0.55	0.46	0.37
90	0.54	0.48	0.39	0.32
Below: AVS constant (30) and f_{FP} and TOC variable				
AVS (µmol/g) = 30	**TOC (%)**			
	0.5	**1**	**3**	**10**
f_{FP} (%)	HQ* value derived when HQ = 1			
10	1.74	1.54	1.28	1.04
50	0.66	0.59	0.49	0.40
70	0.54	0.48	0.40	0.32
90	0.46	0.41	0.34	0.28

New sediments

The new sediments collected for the study had metals concentrations that ranged for near-background for the region to highly contaminated (Table S4), and varying properties (Table 2, main text).

Table S4 New sediment data: concentrations in the sediments (mg/kg).

Sediment	As	Cd	Cr	Cu	Ni	Pb	Zn
1 (Control)	5	0.1	8	20	3	20	47
2	19	0.8	29	35	12	65	200
3	17	1.2	34	63	15	108	358
4	7	1.0	25	108	12	175	**539**
5	16	1.5	39	90	17	152	**517**
6	13	0.2	68	**390**	6	120	230
7 (Ox)	13	0.2	68	**390**	6	120	230
8	6	0.8	23	203	14	**227**	**585**
9 (Ox)	6	0.8	23	203	14	**227**	**585**
10	14	1.9	44	118	20	195	**675**
11	5	1.0	22	198	12	**299**	**442**
12	12	2.7	36	131	16	213	**722**
13	13	2.3	49	145	23	**238**	**833**
14	12	0.4	26	247	12	**345**	**664**
15 (Ox)	12	0.4	26	247	12	**345**	**664**
16	14	0.4	109	**780**	6	100	230
17 (Ox)	14	0.4	109	**780**	6	100	230
18	25	0.9	47	**558**	12	**499**	**936**
19	13	0.3	65	**1050**	7	132	345
20 (Ox)	13	0.3	65	**1050**	7	132	345
21	26	2.8	116	**396**	20	**436**	**1220**
22 (Ox)	26	2.8	116	**396**	20	**436**	**1220**
23	13	**35**	8	65	6	**230**	**1100**
24	20	3.1	72	**365**	28	**626**	**2410**
25	22	2.1	61	**837**	39	**692**	**2680**
26 (Ox)	22	2.1	61	**837**	39	**692**	**2680**
SQGV	20	1.5	80	65	21	50	200
SQG-High	**70**	**10**	**370**	**270**	**52**	**220**	**410**

Seven sediments with (Ox) indicates those that were resuspension-oxidation treated, with the corresponding non-treated sediment being the number above.

SQGV: sediment quality guideline value for protection of 95% marine water species (Simpson et al., 2013b).

SQG-High: high guideline value for sediment quality (Simpson et al., 2013b).

Concentrations are underlined when between SQGV and SQG-High, and **bold** when above SQG-High.

Table S5 Metal concentrations of the overlying water in the static-renewal tests (µg/L).

Sediment	As *	Cd	Cr	Cu	Ni	Pb	Zn
1 (Control)	6.5	0.5	1.0	2.2	1.2	2.5	3.0
2	6.5	0.5	1.0	1.6	2.0	2.5	2.3
3	6.5	0.5	1.0	2.8	2.0	2.5	2.5
4	6.5	0.5	1.0	11.9	3.4	2.5	9.2
5	6.5	0.5	1.0	1.7	2.0	2.5	16
6	6.5	1.0	1.0	5.4	3.9	2.5	264
7 (Ox)	6.5	0.7	1.0	5.9	3.4	2.5	60
8	6.5	0.5	1.0	6.6	3.4	2.5	35
9 (Ox)	6.5	10.7	1.0	1.5	3.4	2.5	245
10	6.5	0.5	1.0	1.7	3.4	2.5	33
11	6.5	0.6	1.0	6.5	1.2	2.5	86
12	6.5	0.5	1.0	1.0	3.4	2.5	8.0
13	6.5	0.5	1.0	4.7	3.4	2.5	44
14	6.5	0.3	0.7	8.7	3.4	0.5	5.8
15 (Ox)	6.5	0.3	0.7	1.8	3.4	0.5	6.3
16	6.5	0.3	0.7	1.1	3.4	0.5	6.2
17 (Ox)	6.5	0.4	0.7	2.3	3.4	0.7	52
18	6.5	0.3	0.7	2.0	3.4	0.5	37
19	6.5	0.4	0.7	15.1	3.4	0.5	9.4
20 (Ox)	6.5	0.3	0.7	9.6	3.4	0.5	7.4
21	6.5	0.3	0.7	24.3	3.4	0.5	23
22 (Ox)	6.5	0.3	0.7	4.5	3.4	0.5	20
23	6.5	0.4	0.7	18.2	3.4	0.5	15
24	6.5	0.3	0.7	1.2	3.4	1.2	7.7
25	6.5	0.3	0.7	1.1	3.4	0.5	19
26 (Ox)	6.5	0.7	0.7	8.9	3.5	0.5	243
WQGV	**12**	**5.5**	**27.4**	**1.3**	**70**	**4.4**	**15**

Seven sediments with (Ox) indicates those that were resuspension-oxidation treated, with the corresponding non-treated sediment being the number above.

WQGV: water quality guideline value for protection of 95% marine water species (ANZG, 2018). The value of 12 µg/L for arsenic (As) is taken from Golding et al. (2022).

The limits of detections (LOD) for ICPAES analyses of the metals in seawater were determined for each batch of samples often varied, being 6.5, 0.3-1.0, 0.7-1.0, 0.5, 1-3.4, 0.5-1.5, 1-2 µg/L for As, Cd, Cr, Cu, Ni, Pb and Zn, respectively.

[*] For arsenic, all the values were <LOD (6.5 µg/L) and the LOD value of 6.5 was used

Table S6 Metal concentrations of the overlying water in the flow-through tests (µg/L).

Sediment	**As ***	**Cd**	**Cr**	**Cu**	**Ni**	**Pb**	**Zn**
1 (Control)	6.5	0.5	1.0	1.7	1.2	2.5	3.0
2	6.5	0.5	1.0	1.3	2.0	2.5	0.5
3	6.5	0.5	1.0	1.3	2.0	2.5	1.5
4	6.5	0.5	1.0	4.5	3.4	2.5	5.3
5	6.5	0.5	1.0	1.3	2.0	2.5	2.1
6	6.5	0.5	1.0	1.0	3.4	2.5	3.4
7 (Ox)	6.5	0.5	1.0	2.0	3.4	2.5	10
8	6.5	0.5	1.0	2.1	3.4	2.5	6.2
9 (Ox)	6.5	0.5	1.0	1.0	3.4	2.5	5.9
10	6.5	0.5	1.0	1.1	3.4	2.5	2.4
11	6.5	0.5	1.0	2.1	1.2	2.5	12
12	6.5	0.5	1.0	1.1	3.4	2.5	1.6
13	6.5	0.5	1.0	4.3	3.4	2.5	62
14	6.5	0.3	0.7	2.0	3.4	0.5	0.5
15 (Ox)	6.5	0.3	0.7	1.4	3.4	0.5	0.5
16	6.5	0.3	0.7	1.8	3.4	0.5	1.0
17 (Ox)	6.5	0.3	0.7	1.9	3.4	0.5	1.4
18	6.5	0.3	0.7	1.2	3.4	0.5	3.2
19	6.5	0.3	0.7	6.1	1.3	0.2	3.8
20 (Ox)	6.5	0.3	0.7	3.1	3.4	0.5	0.8
21	6.5	0.3	0.7	6.3	3.4	0.5	0.5
22 (Ox)	6.5	0.3	0.7	2.1	3.4	0.5	0.8
23	6.5	0.3	0.7	2.0	3.4	0.5	2.1
24	6.5	0.3	0.7	2.1	3.4	0.5	1.4
25	6.5	0.3	0.7	3.2	3.4	0.5	2.7
26 (Ox)	6.5	0.3	0.8	5.9	3.4	0.5	16
WQGV	**12**	**5.5**	**27.4**	**1.3**	**70**	**4.4**	**15**

Seven sediments with (Ox) indicates those that were resuspension-oxidation treated, with the corresponding non-treated sediment being the number above.

WQGV: water quality guideline value for protection of 95% marine water species (ANZG, 2018). The value of 12 µg/L for arsenic (As) is taken from Golding et al. (2022).

The limits of detections (LOD) for ICPAES analyses of the metals in seawater were determined for each batch of samples often varied, being 6.5, 0.3-1.0, 0.7-1.0, 0.5, 1-3.4, 0.5-1.5, 1-2 µg/L for As, Cd, Cr, Cu, Ni, Pb and Zn, respectively.

* For arsenic, all the values were <LOD (6.5 µg/L) and the LOD value of 6.5 was used

Table S7 Survival and reproduction (percentage to control) of *Melita plumulosa* in sediment toxicity tests.

Sediment	Static-renewal				Flow-through			
	Survival (%)		Reproduction (%)		Survival (%)		Reproduction (%)	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1 (Control)	100	5	100	9	100	5	100	13
2	105	5	101	4	93	4	70	5
3	95	6	71	12	102	2	57	8
4	91	4	36	7	89	9	66	8
5	102	4	58	13	85	8	65	6
6	85	5	34	4	86	25	65	31
7 (Ox)	100	2	42	17	89	6	76	15
8	104	2	23	8	102	3	64	9
9 (Ox)	93	4	34	8	105	5	40	3
10	80	4	50	11	92	8	62	5
11	82	0	22	6	93	2	60	4
12	83	4	46	10	85	15	77	6
13	100	2	35	5	100	6	46	9
14	102	7	62	4	102	3	71	16
15 (Ox)	96	4	71	8	98	4	89	4
16	102	4	78	7	95	6	104	11
17 (Ox)	98	4	65	7	96	4	82	6
18	83	4	46	12	102	3	95	6
19	87	4	17	4	107	6	73	7
20 (Ox)	98	2	48	8	100	12	82	8
21	58	11	10	4	105	2	33	10
22 (Ox)	102	2	58	13	95	8	114	15
23	100	3	30	9	93	4	48	17
24	96	6	68	5	84	8	105	15
25	95	4	68	9	90	7	47	5
26 (Ox)	100	2	29	8	107	5	46	8

Seven sediments with (Ox) indicates those that were resuspension-oxidation treated, with the corresponding non-treated sediment being the number above.

All data are % of control, based on 4 replicates.

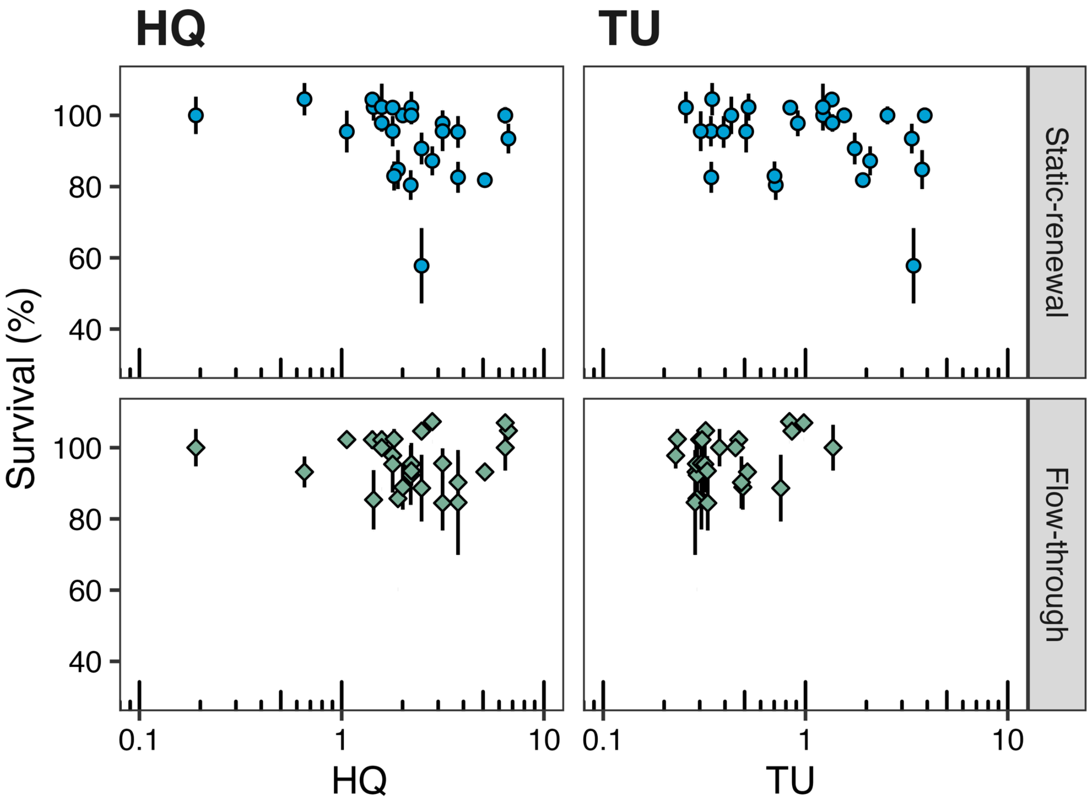


Figure S3. Survival of *Melita plumulosa* in the static-renewal and the flow-through sediment tests. HQ is the hazardous quotient based on the metal concentrations in the sediment, and TU is the toxic unit based on the metal concentrations in the overlying water.

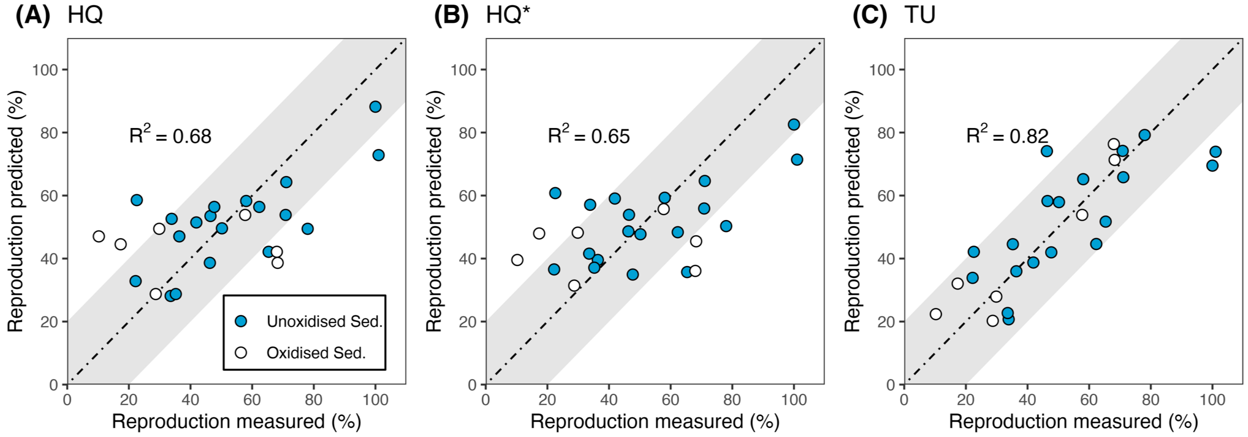


Figure S4. Measured and predicted reproduction for the new static-renewal data set (N = 26) using different models: (A) the HQ-based model (Eq. 5), (B) the HQ*-based model (Eq. 6), and (C) the TU-based model (Eq. 7). The open symbols indicate the 7 oxidised sediments (Table 2, Table S3). The dotdash line represents the 1:1 line, and the band represents a factor of two.

Table S8 Model prediction of Categories of toxicity, N = 91 static-renewal test results

S8A Toxicity categories evaluated in comparison of measured with predicted

	Category	Amphipod reproduction
1	**No toxicity**	>80% reproduction at day 10 termination
2	**Low**	>60-80% reproduction at day 10 termination
3	**Moderate**	>40-60% reproduction at day 10 termination
4	**High**	>20-40% reproduction at day 10 termination
5	**Very high**	0-20% reproduction at day 10 termination

S8B Comparison of measured with predicted results for N = 91 test (Categories / Accuracy-count)

Category	Count for Measured Category N, (% per Category)	Prediction: Category Accuracy											
		HQ Model						HQ* Model					
		More toxic		Correct	Less toxic			More toxic		Correct	Less toxic		
		-2	-1	0	+1	+2	+3	-2	-1	0	+1	+2	+3
No toxicity	41 (45%)	1	15	25	0	0	0	0	5	36	0	0	0
Low	21 (23%)	0	4	12	5	0	0	3	5	4	9	0	0
Moderate	12 (13%)	0	0	4	7	1	0	1	1	5	5	0	0
High	12 (13%)	0	0	0	4	8	0	0	0	6	4	1	1
Very high	5 (5%)	0	0	0	1	1	3	0	0	0	4	1	0
Total (n)	**91**	**1**	**19**	**41**	**17**	**10**	**3**	**4**	**11**	**51**	**22**	**2**	**1**
Overall Accuracy	**100%**	**1%**	**21%**	**45%**	**19%**	**11%**	**3%**	**4%**	**12%**	**56%**	**24%**	**2%**	**1%**

The “**Count**” for Measured Category is the number of Test results within each Toxicity Category. “**Prediction: Category Accuracy**” (5 columns for each of HQ and HQ*) partitions the comparison of the prediction with measured toxicity by accuracy (the Count distributed): Correct (i.e., measured and predicted are same category), and More toxic by 1 or 2 categories, and Less toxic by 1, 2, 3 categories.

As examples:
(a) if the Measured test result was High and the Prediction was Low the comparison would be counted as a +2 result (Predicted Less toxic by two categories);
(b) if the Measured test result was No toxicity and the Prediction was Low the comparison would be counted as a -1 result (Predicted More toxic by one category)

The **Overall Accuracy** (bottom row) is the sum of each ‘Accuracy’ column divided by the total number of results, as %.

S8C Presentation of S8B results as a percent of the Count for each Toxicity Category, and in bottom row for all data (N = 91) distributed by Accuracy.

Category	Count for Measured Category N	HQ Model Accuracy %						HQ* Model Accuracy %					
		More toxic		Correct	Less toxic			More toxic		Correct	Less toxic		
		-2	-1	0	+1	+2	+3	-2	-1	0	+1	+2	+3
No toxicity	41	2%	37%	61%	0%	0%	0%	0%	12%	88%	0%	0%	0%
Low	21	0%	19%	57%	24%	0%	0%	14%	24%	19%	43%	0%	0%
Moderate	12	0%	0%	33%	58%	8%	0%	8%	8%	42%	42%	0%	0%
High	12	0%	0%	0%	33%	67%	0%	0%	0%	50%	33%	8%	8%
Very high	5	0%	0%	0%	20%	20%	60%	0%	0%	0%	80%	20%	0%

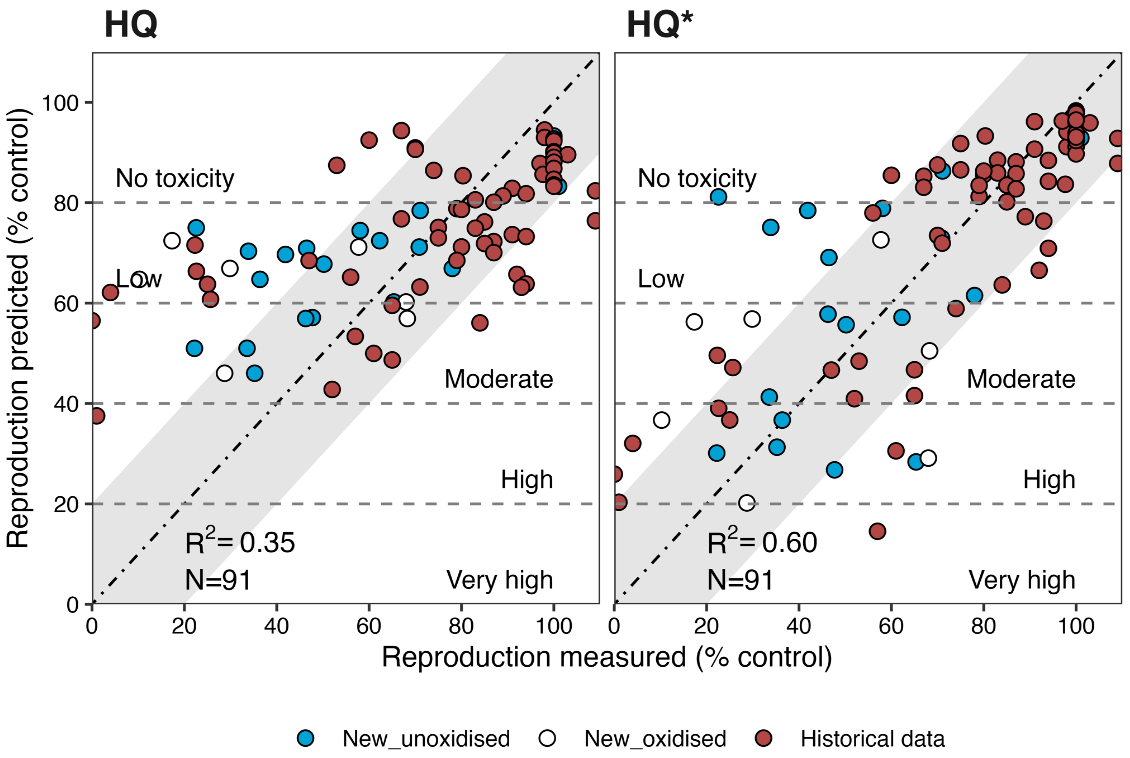


Figure S5 Measured and predicted amphipod reproduction comparison for the two models: (A) HQ (Eq. 4) and (B) HQ* (Eq. 5) for the combined static-renewal sediment toxicity data set. Arbitrary Categories of Toxicity of 20% are designated by the horizontal dashed lines. The dotdash line represents the 1:1 line, and the band represents a factor of two.

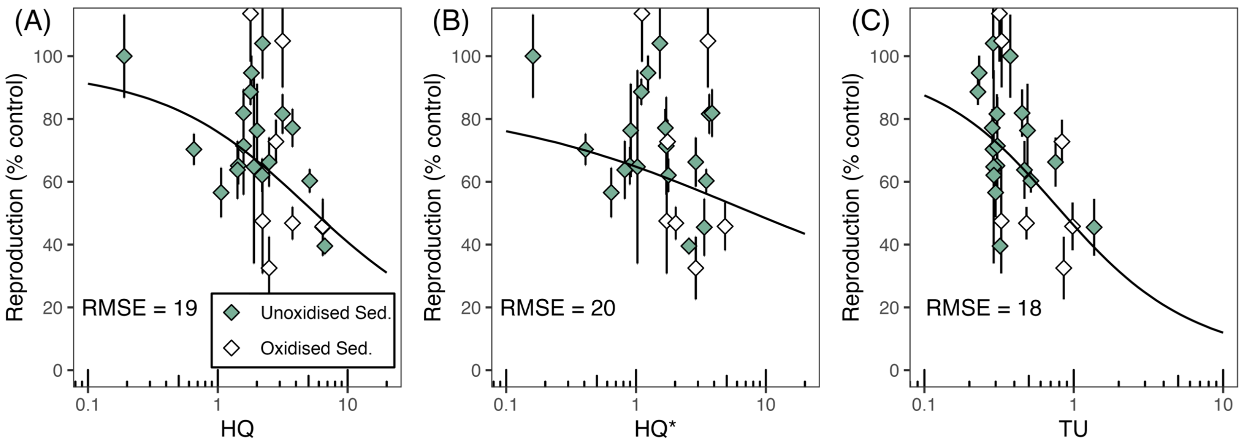


Figure S6 The logistic relationship between reproduction of amphipod and the three models: (A) HQ (Eq. 4), (B) HQ* (Eq. 5), and (C) TU (Eq. 7) for the new flow-through sediment toxicity data (N = 26).