

University of Windsor

Scholarship at UWindsor

Great Lakes Institute for Environmental
Research Publications

Great Lakes Institute for Environmental
Research

3-26-2016

Assessment of hazard metrics for predicting field benthic invertebrate toxicity in the Detroit River, Ontario, Canada

Kerry N. Mcphedren
University of Saskatchewan


Alice Grigicak-Mannion
University of Windsor

Gord Paterson
State University of New York, Syracuse

Briggs Tedd
Ontario Ministry of Environment and Climate Change

Jan J.H. Ciborowski
University of Windsor

Below this page find additional works that <https://scholar.uwindsor.ca/glierpub>

 Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#), and the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Mcphedren, Kerry N.; Grigicak-Mannion, Alice; Paterson, Gord; Tedd, Briggs; Ciborowski, Jan J.H.; Haffner, G. Douglass; and Drouillard, Ken G.. (2016). Assessment of hazard metrics for predicting field benthic invertebrate toxicity in the Detroit River, Ontario, Canada. *Integrated Environmental Assessment and Management*.
<https://scholar.uwindsor.ca/glierpub/84>

This Article is brought to you for free and open access by the Great Lakes Institute for Environmental Research at Scholarship at UWindsor. It has been accepted for inclusion in Great Lakes Institute for Environmental Research Publications by an authorized administrator of Scholarship at UWindsor. For more information, please contact scholarship@uwindsor.ca.

Authors

Kerry N. Mcphedren, Alice Grigicak-Mannion, Gord Paterson, Briggs Tedd, Jan J.H. Ciborowski, G. Douglass Haffner, and Ken G. Drouillard

Revised March 12, 2016

Title: Assessment of hazard metrics for predicting field benthic invertebrate toxicity in the Detroit River, Ontario, Canada

Accepted by *Integrated Environmental Assessment and Management*, March, 2016.

Authors:

Kerry N. McPhedran[†], Alice Grgicak-Mannion[‡], Gord Paterson[§], Ted Briggs^l, Jan J.H. Ciborowski[#], G. Douglas Haffner[‡] and Ken G. Drouillard^{‡,*}

[†] College of Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, Saskatchewan, S7N 5A9, Canada

[‡] Great Lakes Institute for Environmental Research, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4, Canada

[§] College of Environmental Science and Forestry, State University of New York, 1 Forestry Drive, Syracuse, New York, 13210, United States

^l Ontario Ministry of the Environment, 733 Exeter Road, London, Ontario, N6E 1L3, Canada

[#] Department of Biological Sciences, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4

* To whom correspondence may be addressed. T: +1 519 253 3000; F: +1 519 971 3686; E: kgd@uwindsor.ca

Revised March 12, 2016

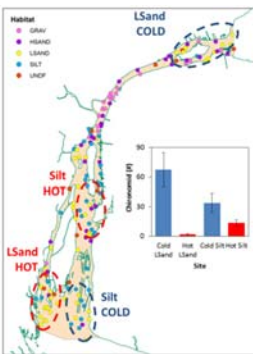
ABSTRACT

Numerical sediment quality guidelines (SQGs) are frequently used to interpret site specific sediment chemistry and predict potential toxicity to benthic communities. These SQGs are useful for a screening line of evidence (LOE) that can be combined with other LOEs in a full weight of evidence (WOE) assessment of impacted sites. Three common multi-chemical hazard quotient methods (PEC- Q_{avg} , PEC- Q_{met} and PEC- Q_{sum}) and a novel (Hazard Score; HZD) approach were used in conjunction with a consensus based set of SQGs to evaluate the ability of different scoring metrics to predict the biological effects of sediment contamination under field conditions. Multivariate analyses were first used to categorize river sediments into distinct habitats based on a set of physicochemical parameters to include gravel, low and high flow sand and silt. For high flow sand and gravel, no significant dose-response relationships between numerically dominant species and various toxicity metric scores were observed. Significant dose-response relationships were observed for Chironomid abundances and toxicity scores in low flow sand and silt habitats. For silt habitats the HZD scoring metric provided the best predictor of Chironomid abundances compared to various PEC-Q methods according to goodness of fit tests. For low flow sand habitats, PEC- Q_{sum} followed by HZD, provided the best predictors of Chironomid abundance. Differences in apparent Chironomid toxicity between the two habitats suggest habitat specific differences in chemical bioavailability and indicator taxa sensitivity. Using an index of biological integrity (IBI) method, the HZD, PEC- Q_{avg} and PEC- Q_{met} approaches provided reasonable correlations with calculated IBI values in both silt and low flow sand habitats but not for gravel or high flow sands. Computation differences between the various multi-chemical toxicity scoring metrics and how this contributes to bias in different estimates of chemical mixture toxicity scores are discussed and compared.

Revised March 12, 2016

KEYWORDS: hazard score, sediment quality guidelines (SQGs), benthic community impacts, connecting channel

JOURNAL COVER IMAGE



Revised March 12, 2016

INTRODUCTION

Multiple lines of evidence (LOEs) are commonly used to assess the impact of sediment contamination on benthic organisms in North America and worldwide (Long et al. 2006). Typical LOEs include bulk sediment chemistry, toxicity tests, bioaccumulation tests and benthic community composition (Chapman and Anderson 2005). Within the LOE approach, numerical sediment quality guidelines (SQGs) are used as screening tools to assist in interpreting sediment chemistry data (Long et al. 2006; Ritter et al. 2011). By using multiple LOEs, a weight of evidence (WOE) assessment of site contamination can be useful for a complete understanding of sediment contaminant impacts. Although SQGs are typically assigned according to jurisdictions, inter-jurisdictional SQGs exhibit wide ranges of chemical concentrations that can span two to three orders of magnitude in their predicted dose-response for a given chemical (MacDonald et al. 2000). SQGs are most commonly developed and assessed using standardized sediment toxicity assays (e.g. amphipod toxicity, Microtox[®] bioluminescence, sea urchin fertilization tests) which are meant to serve as surrogates for actual benthic species (Wenning et al. 2004). Given that these tests are performed under controlled laboratory conditions, it is difficult to determine how accurately the SQGs can predict toxicity under more complex field conditions, which include a diversity of environmental (e.g., site specific differences in chemical bioavailability) and ecological attributes as well as complex chemical mixture effects (Bay and Weisberg 2010; Chapman 1996). Given that the purpose of SQGs is to provide a screening level assessment of *in situ* toxicity, it is imperative that attempts at field assessment of SQGs be performed on a regular basis and cover multiple environment types (Ritter et al. 2011; Bay et al. 2012).

Revised March 12, 2016

In order to establish a proper methodology for the field assessment of a set of SQGs, both environment (habitat) and chemical mixture interactions need to be taken into consideration. The effect of chemical exposures on zoobenthic communities can be obscured by the overriding habitat-specificity of many aquatic invertebrates, especially in lotic systems (Wright 1995; Parsons and Norris 1996; Reynoldson et al. 1997; Reynoldson et al. 2001). For example, in a review of SQGs, Long et al. (2006) indicated the need for increased understanding of habitat-specific local features such as particle size, depth and velocity in addition to chemical-specific stressors on the benthic community. Habitat classification is more commonly considered in surveys designed to assess benthic community structure such as with the reference condition approach (RCA). These approaches provides a multivariate means of matching habitat attributes in a sample test location against a library of reference sites prior to evaluating for differences in benthic community composition at the test location (Wright 1995; Parsons and Norris 1996; Reynoldson et al. 1997). Presumably comparable approaches to habitat-matching could be adopted under a field SQG assessment, although the use of such a strategy seems to be uncommon. In such a case, dose-response relationships in the numerical presence of a habitat-appropriate indicator species or community composition can be evaluated across a gradient of chemical contamination measured within a given habitat type. Applying such a methodology for field SQG assessment may also have the advantage of being able to examine for differences in the performance of SQGs between habitat types and in the identification of site specific indicator species.

A second complication related to field assessment of SQGs involves the ability to define contamination gradients owing to chemical mixture effects arising from differences in the sources and environmental distribution of individual chemicals. These chemical-specific

Revised March 12, 2016

differences can lead to disparate patterns in contamination where presumably ‘clean’ conditions for one chemical of study may be highly degraded for another. There exist several hazard quotient approaches currently available that combine SQGs and site specific contaminant measurements for multiple chemicals to compute a single toxicity score. However, all toxicity score computation methods incorporate sets of assumptions that can introduce anomalies that bias toxicity assessment towards over- or under- estimation of the actual site specific toxicity. For example, most studies adopting a hazard quotient approach consider deviations in site-specific chemical residues with SQGs attributed to high biological impact (e.g. the severe effect concentration (SEC) or probable effect concentration (PEC)) but ignore SQGs for the same chemicals that are associated with more conservative estimates of toxicity (e.g. threshold effect concentration (TEC) or low effect concentration (LEC)). Thus, hazard quotient approaches by default assume a linear dose/response relationship that is based on one calibration point even though toxicity related to animal mortality often takes on a sigmoidal distribution which necessitate multiple calibration points (Ritter et al. 2011; MacDonald et al. 2000).

Similarly, toxicity score metrics differ in their computation schemes that make the method more or less sensitive to the number of chemicals included and/or diversity of chemical signatures generated in a given survey. For example, different toxicity score metrics consider the sum, average or weighted average hazard quotients for sets of priority contaminants. Toxicity scores generated as the sum of hazard quotients can be biased to over-predicting toxicity when many chemicals are included in the metric even when all of the chemicals are below the SQGs at a given location. Metrics generated by an average hazard quotient approach can lead to underestimates of toxicity when several chemicals have very low site specific concentrations. These low concentration chemicals can effectively dilute the toxicity score value even when one

Revised March 12, 2016

or more contaminant concentrations exceed their SQGs for high biological impact. One of the objectives of this study was to introduce an alternative toxicity score metric, henceforth identified as the hazard score (HZD) metric, which assumes a sigmoidal toxicity distribution, adopts a multi-point calibration scheme based on existing SQGs and establishes a threshold effect concentration, below which chemicals included in the metric have no influence on the summed toxicity score. This new toxicity metric is then compared with more classic multi-chemical hazard quotient (designated as PEC-quotient or PEC-Q) approaches to determine, which metric if any, provides the best prediction of field toxicity as evaluated using field assessment data.

Field assessment data used for this study were generated from a combined high resolution (n=150 sampling stations) sediment chemistry and benthic community assessment survey conducted for the Detroit River. First, a habitat classification scheme was developed that used a multivariate approach to designate distinct sediment habitat types within the Detroit River survey locations using habitat variables previously described as being important to benthic community structure. Following this, the benthic community database was used to identify potential benthic invertebrate indicators within each habitat type by choosing organisms that demonstrated high association, high abundance (but also high variation in abundance) across sample locations of a given habitat. The relative abundance of each benthic invertebrate indicator was then used as a surrogate measure of 'field toxicity' and compared with the different SQG-based toxicity score metrics. In addition, a multi-metric index of biological integrity (IBI) was developed for each unique habitat type of the Detroit River according to the methods described by Reynoldson et al. (1997). The habitat specific IBIs were also contrasted with each of the SQG-based toxicity scores to determine which, if any, showed the strongest correlations. This permitted a set of

Revised March 12, 2016

independent contrasts between different chemical mixture assessment approaches (SQG-based toxicity scores), indicator species abundance and multi-metric community composition changes.

METHODS

Area of Interest Background

The Detroit River is a 54 km connecting channel divided between the State of Michigan (United States) and the Province of Ontario (Canada). The sample collection protocol for the current study has been described previously by Drouillard et al. (2006). Briefly, 150 sampling sites were identified based on a stratified random design within three reaches (upper, middle and lower) of the Detroit River (Fig. 1). Surface sediments were collected using a petite Ponar grab sampler, whereupon multiple grabs were used to provide a standardized total volume of 2 L of sediment. Water depth was measured on site. Bottom water velocity was estimated from a previously developed 3-D hydrological model (Reitsma et al. 2003). Sediment samples were analysed for physical (total organic carbon (TOC), silt, sand and gravel), chemical (arsenic, cadmium, chromium, copper, iron, lead, mercury, nickel, zinc, hexachlorobenzene (HCB), dichlorodiphenyldichloroethylene (DDE), total polychlorinated biphenyls (PCBs) and total polyaromatic hydrocarbons (PAHs)) and benthic community data (Table 1). Complete physicochemical analytical methods are described by Drouillard et al. (2006) and Szalinska et al. (2006). A second 2 L sediment sample from each sample location was collected by similar methods. Samples were coarsely sieved (600 μm) at the sampling location to remove most of the fine materials and the contents of the sieve bucket emptied into a plastic bag and preserved with Kahles solution. At the laboratory, zoobenthos were sorted and identified to common lowest taxonomic ranks under a dissecting scope and stored in glass vials containing 80% ethanol. For

Revised March 12, 2016

quality assurance, 10% of samples were resorted and compared with the previous assessment. Overall, resamples agreed with originals within 4% error.

Toxicity Metric Scoring Approaches

The Michigan Department of Environmental Quality (MDEQ) SQGs are consensus-based values as reviewed and recommended by Macdonald et al. (2000). These SQGs provide values analogous to the threshold effects concentration (TEC) and probable effects concentration (PEC; Table S1). The SQG-based toxicity score approaches considered in this study include three variations of previously developed multi-chemical PEC-quotient approaches and a new Hazard score approach described below.

PEC-Quotient Approaches

For each chemical where an SQG is available and sediment concentration is reported, the PEC- Q_x is calculated as:

$$PEC - Q_x = \frac{c_x}{PEC_x} \quad (1)$$

where c_x is the measured chemical concentration ($\mu\text{g/g}$ dry weight) in a sediment sample and PEC_x is the probable effect concentration ($\mu\text{g/g}$ dry weight) for the chemical of study based on the SQG. The multi-chemical PEC-quotient metrics all use chemical and sediment specific PEC- Q_x values as specified in Eq. 1 but differ in their computation method. Three previously published multi-chemical PEC-quotient (PEC-Q) approaches were adopted and include the average (PEC- Q_{avg}), a weighted average PEC-Q with metals as a single group (PEC- Q_{met}) and the sum (PEC- Q_{sum}).

Revised March 12, 2016

The $PEC-Q_{avg}$ provides the mean $PEC-Q$ value across all chemicals measured (n is the total number of chemicals where $PEC-Q$ values were derived) in a given sediment sample as reviewed by Long et al. (2006). The $PEC-Q_{avg}$ is calculated according to:

$$PEC - Q_{avg} = \frac{\sum_{i=1}^n PECQ_x}{n} \quad (2)$$

The $PEC-Q_{met}$ is a derivation of the $PEC-Q$ metric considered by Ingersoll et al. (2001) and modified herein. In their approach, Ingersoll et al. (2001) weighted a $PEC-Q_{avg}$ generated for all metals in the sample with $PEC-Q_x$ values for total PAHs and sum PCBs. The current derivation includes HCB and DDE using the same assumption and the consideration that both HCB and DDE have available SQG criteria and were ubiquitous within the Detroit River. In the present study, $PEC-Q_{met}$ is calculated according to:

$$PEC - Q_{met} = \frac{\sum_{i=1}^{n_{met}} \frac{PECQ_x(metals)}{n_{met}} + PECQ_{HCB} + PECQ_{DDE} + PECQ_{PCBs} + PECQ_{PAHs}}{5} \quad (3)$$

Where n_{met} is the number of metals identified in the sample, $PEC-Q_{x(metals)}$ is the $PEC-Q_x$ generated for each metal analyzed in the sample, $PEC-Q_{HCB}$, $PEC-Q_{DDE}$, $PEC-Q_{PCBs}$ and $PEC-Q_{PAHs}$ are the $PEC-Q_x$ determined for HCB, DDE, total PCBs and total PAHs, respectively. Commonly, $PEC-Q_{met}$ generates a toxicity score which is greater than the $PEC-Q_{avg}$ although exceptions can occur depending on the relative magnitude of $PEC-Q_x$ for individual metals in the sample.

The $PEC-Q_{sum}$ is the sum of $PEC-Q_x$ across chemicals of study and is defined by:

$$PECQ_{sum} = \sum_{i=1}^n PECQ_x \quad (4)$$

$PEC-Q_{sum}$ always produces a toxicity score greater than $PEC-Q_{avg}$ and $PEC-Q_{met}$ and is the most sensitive index with respect to the number of chemicals used to generate the toxicity score.

Revised March 12, 2016

By convention, PEC-Q toxicity scores exceeding a value of 1 are considered toxic, or having the potential to be toxic. In order to standardize this score with the field toxicity determined in sediment samples it was assumed that a PEC-Q value of 1 is equivalent to 50% toxicity. In other words, this assumes that the PEC-SQG establishes a probable effect level corresponding to toxicity similar to an LC₅₀ value for an indicator species present within the sample. One other modification was that each of the PEC-Q toxicity scores was capped at a value of 100% toxicity (i.e., PEC-Q values greater than 2 were set to a value of 2 and 100% toxicity, respectively). This yields a toxicity score for each PEC-Q method that ranges from 0 to 100% toxicity. Overall, it was expected that PEC-Q_{avg} would produce the lowest estimate of toxicity, PEC-Q_{met} an intermediate toxicity estimate and PEC-Q_{sum} would yield the highest toxicity estimate for individual sampling sites.

Hazard Score (HZD) Approach

In this alternative toxicity score metric, both the TEC and PEC SQG values are used to generate chemical specific sigmoidal dose-response toxicity curves according to a 2 point calibration. The dose response curve is then used to calculate a toxicity value for each chemical present in the sample based on measured chemical concentrations in sample. The Hazard Score (HZD) is the sum of %toxicity values generated across chemicals similar to the PEC-Q_{sum} method with modification to establish a 0% toxicity value where sediment concentrations are less than TEC and at the upper range to cap toxicity values at 100% when sum %toxicity exceeds 100.

Revised March 12, 2016

Here the TEC is assigned a 5% toxicity level and the PEC was initially set to 50% toxicity. This is the equivalent of establishing SQG TEC and PEC as chemical specific LC₅ and LC₅₀ values, respectively. A sigmoidal dose/response toxicity curve is generated according to:

$$Effect(\%) = \frac{100}{1+(A \cdot e^{-kC})} \quad (5)$$

where Effect(%) is the anticipated toxicity (ranging from 0 to 100%), A is a constant that determines the curvature of the dose-response curve, k is the chemical-specific toxicity coefficient and C is the measured sediment chemical concentration. The values of A and k are iteratively solved for each chemical such that the 5% and 50% toxicity predictions conform to TEC and PEC concentration values (see Supplementary Information Table S1). Fitted curves to Eq. 5 for each chemical are included in Supplementary Information, Figure S1.

The HZD is the sum of each chemical Effect(%) for a specific sampling station. Two modifications were made to the computational algorithm of HZD. For any given chemical, when the measured sediment concentrations was less than the TEC, its Effect(%) value was set to zero. This modification was performed because Eq 5 generates a non-zero intercept and has a predicted toxicity (above 0 and less than 5%) when the sediment concentration for the contaminant has a zero concentration. By setting all concentrations <TEC to a 0 Effect(%), the cumulative effect of multiple chemicals with very low or zero concentration values are removed from the score. The second modification was similar to that performed for PEC-Q approaches, i.e. the maximum HZD value was capped at 100%, thus when sum Effect(%) >100, the HZD was set to 100. Based on these modifications, HZD toxicity scores can range from 0 to 100% for a given sampling location.

Revised March 12, 2016

Habitat and Benthic Invertebrate Indicator Species Identification

The suite of physicochemical environmental variables measured at each site are included in Table 1. A principal component analysis (PCA) was performed on sediment variables deemed as habitat attributes using log-transformed data and applying a correlation matrix. The PCA identified depth (m), velocity (m/s), TOC (%), gravel (%), sand (%) and silt (%) as significant parameters (i.e. each having loadings >0.7 on the first 2 PCA axes) contributing to the ordination of samples. PC1 exhibited strong loadings for habitat attributes associated with depth, water velocity and gravel with TOC, sand and silt being strongly loaded on PC2 (See Supplementary Information, Table S2). Individual sample scores on PC1 and PC2 were used to further characterize sample sites with preliminary user-identified ellipses constructed around sites that appeared to exhibit similar habitat groupings (Fig. 2A). Habitat classifications assigned from the results of the initial PCA (Fig. 2A) were verified by discriminant function analysis (DFA) which statistically tests whether each site has been appropriately categorized based on the habitat attribute variables (Fig. 2B; Supplementary Information, Table S3). All sites grouped into a given habitat cluster, and whose assignment was verified by DFA, were then categorized as the same habitat type. Some sites that had intermediate characteristics across habitats and could not be assigned by DFA were excluded from further analysis. This habitat classification scheme assumes that the habitat attributes identified in the ordination scheme are appropriate habitat descriptors for benthic invertebrate indicators and that different sampling sites of the same habitat type all have the same potential to support indicator species survival and benthic invertebrate communities except as modified by degree of chemical contamination.

For the overall benthic species composition, a second PCA was used to reduce the total amount of species groups and to determine strongly covarying species assemblages (See

Revised March 12, 2016

Supplementary Information, Table S4). Individual species representing $\leq 0.5\%$ of total abundances were not included in the PCA as these individuals generally do not significantly increase the total variance explained in the dataset (Reynoldson et al. 1995). Zoobenthos abundances were $\log(x+1)$ transformed prior to PCA analysis. Indicator species for each habitat type were evaluated and chosen on the basis of the above benthic species composition PCA. The numerically dominant species showing strong loadings on each PCA axis were selected for further consideration as prospective indicators. For each of these species, box and whisker charts were prepared to identify which habitat the species was most associated with and where large ranges in abundance values were apparent (See Supplementary Information Fig. S2). Habitat specific indicator species was then chosen for each habitat type on the basis of the above box and whisker charts.

All multivariate analyses performed to classify sediments and benthos was completed using SYSTAT version 8.0 for Windows.

Field Toxicity Estimates

The field toxicity determined for a given sampling station was used to assess toxicity predictions generated by the various PEC-Q metrics or the HZD metric. Field toxicity values for each sediment site were generated based on the relative abundance of a benthic indicator species (or taxa) present in the sample contrasted against the mean upper abundance of the same species present in other samples of the same habitat designation.

Field toxicity scores were generated for each indicator species and each site within a given habitat type. The number of the indicator species present in a standard 2 L sediment sample at a given site (A_x) was divided by the average of the five highest abundance values (A_{high})

Revised March 12, 2016

determined for that species across all sites within the same habitat. The relative abundance metric was then converted into a toxicity score according to:

$$Toxicity(\%) = \left(1 - \frac{Ax}{A_{high}}\right) * 100\% \quad (6)$$

A toxicity of 100% was assigned when the indicator species was not present in the sample.

Comparisons between SQG-generated toxicity scores and field toxicity scores were performed using goodness to fit tests of predicted toxicity (y-axis) against field toxicity (x-axis). Goodness of fit test results were contrasted between different toxicity score metrics by evaluating the coefficient of determination of the regression, determining whether the slope differed from a value of unity and the intercept differed from a value of one. Goodness of fit tests forcing the intercept to zero were also performed and evaluated against different toxicity metrics by comparing both the R^2 value and proximity of the slope to a value of unity.

Index of Biological Integrity (IBI) Assessment Method

The multimetric approach scoring methodology was adapted from Reynoldson et al. (1997) without consideration of ‘reference-condition’ sites which were not available in the current study. Metrics considered included: (1) total abundance of organisms counted; (2) number of families identified; (3) %EPT (Ephemeroptera, Trichoptera, and Plecoptera); (4) %Chironomids; (5) number of EPT taxa; (6) % dominance; (7) Shannon-Wiener Index; and (8) Evenness. Each of the above metrics were assessed based on the 25th and 75th percentiles of scores generated across all samples for a given habitat type and the calculated values are shown in Table S5. Site specific scoring for each metric was based on the following: score values <25th percentile within a habitat given as 1; values between 25th and 75th percentile were given a score of 3; values >75th percentile given a score of 5. All metrics had a positive relationship between the metric value

Revised March 12, 2016

and the scoring value (i.e., a score of 5 was given to the highest metric values) other than for the %dominance which was negatively related to scoring value (i.e., a score of 5 was given to the lowest metric values). Overall, the eight metrics used resulted in a minimum value of eight and a maximum score value of 40. These IBI values were used analogously to the Actual(%) toxicity assessments as in the preceding section 'Field Toxicity Estimates'.

RESULTS AND DISCUSSION

Habitat

Principal components analysis was applied to the habitat variables: depth (m), velocity (m/s), TOC (%), gravel (%), sand (%) and silt (%) to account for inter-correlations between variables and reduce the dimensionality of the dataset. The first two components of the PCA were found to cumulatively explain 78.5% of the variation in the data (Table S2). Depth, velocity and gravel were strongly associated with the PC1 axis, and TOC, sand and silt were associated with the PC2 axis. No variables showed strong associations with PCA axes 3, 4 and 5 and given that these axes contributed limited amounts to the variation in the data, they were ignored in habitat characterization. After examining the clustering of sites across the first two PCA axes, four main habitat types were identified consisting of silt (Silt), low velocity sand (LSand), high velocity sand (HSand) and gravel (Grav) (Fig. 2A). Initially, stations were clustered into the habitat types by manually drawing ellipses around similarly ordinated stations as defined by their scores on the PC1 and PC2 axes. DFA was subsequently used to test the assignment of habitat type for each station into the four major habitat types. Initial DFA indicated 85% of sites were correctly classified within the user-identified habitats. After reclassification of improperly classified sites, a second DFA indicated 95% of sites were correctly identified (Table S3). All sites identified by

Revised March 12, 2016

habitat via DFA were used for further analysis (Fig. 2B). Among the 136 stations included in the analysis, 124 stations could be assigned to a habitat type with a 95% confidence. The remaining stations, of unknown or mixed habitat association, were censored from further analysis. Each of the Silt, LSand, HSand and Grav habitats had 33, 45, 28 and 18 total sites, respectively. Generally, habitats were distributed as expected, with the gravel/high flow sand areas found within the upper reach and silt/low flow sand areas generally within the lower reach or near depositional zones of islands (Fig. 1).

An important step in the determination of benthic community impairment is the identification of groups of habitats with similar characteristics (Long et al. 2006). Of the 15 measured environmental variables, only six were needed to delineate the four distinct habitats: depth, velocity, TOC and grain sizes (gravel, sand and silt) (Table 1). Previous studies have indicated the importance of each of these parameters in controlling benthic community composition (Reynoldson et al. 1995; Reynoldson et al. 1997; Rae 1985). Apart from the presently measured parameters, previous researchers have found latitude/longitude, total phosphorus, total nitrogen, water pH and alkalinity to be useful predictors for habitat types (Reynoldson et al. 1995; Reynoldson et al. 1997). The latter were not included in the present analysis.

Benthic Community

The results of PCA for zoobenthos communities indicated the presence of three distinct groups: Group 1 includes Amphipoda, Hydrozoa, Turbellaria and Gastropoda; Group 2 includes Chironomidae and Hexagenia; and Group 3 includes Oligochaeta (Table S4). Group 3 species (Oligochaeta) were completely dominant in areas of the Detroit River with high organic contaminant contamination (Farara and Burt 1993). Previously, Oligochaetes were especially

Revised March 12, 2016

dominant (> 90%) in the highly-polluted Trenton Channel within the Detroit River (Besser et al. 1996).

For each of the three identified groups, the most numerically dominant species were Amphipoda (Group 1), Chironomidae (Group 2) and Oligochaeta (Group 3). Box and whisker plots of group abundances for each habitat (Grav, HSand, LSand and Silt) are shown in Supplementary Information (Figure S2). Both the Chironomidae and Oligochaeta abundances were highest in the LSand and Silt habitats and lowest in the Grav and HSand habitats. In contrast, the Amphipoda were highest in the Grav and HSand environments and lowest in the LSand and Silt habitats. Given absolute abundances and patterns in the data, the following indicator species were identified for each habitat type: Oligochaetes for all four habitat types given their general abundance in each environment; Chironomidae for LSand and Silt; and Amphipoda for Grav and HSand.

Goodness of Fit Tests

Given the above defined indicator groups (Amphipoda, Chironomidae and Oligochaeta) the species specific field toxicities (Toxicity %) were plotted against the various scoring approaches for each habitat type with results shown in Table 2. There was no relationship between the field toxicity and SQG-generated estimate of toxicity for the Amphipods for the Grav ($R^2 = 0.00$ to 0.01) or HSand ($R^2 = 0.00$ for all) habitats for any of the four scoring metrics (see Supplementary Information Fig. S3 for a presentation of field toxicity vs HZD score for these groups). The Oligochaeta were generally abundant across sites for all habitats (Fig. S2). Interestingly, the Oligochaeta showed negative correlations between field and SQG-metric generated toxicity scores for three of the four habitats evaluated including Silt (PEC- Q_{avg} and

Revised March 12, 2016

PEC-Q_{met}), LSand (PEC-Q_{avg} and PEC-Q_{met}) and HSand (all metrics) (Table 2). The IBI values showed no correlations with SQG scores for the HSand (Fig. S4 and Table 3) and Grav (Fig. S5 and Table 3) habitats. Thus assessment of biological impact (either by indicator species abundance or computed IBI metrics) in Grav and HSand as they relate to presence of sediment associated priority pollutants in the Detroit River remains inconclusive.

The Chironomidae were abundant in Silt and LSand environments. For this indicator, toxicity was correlated to all toxicity scoring methods as examined in more detail below. The goodness of fit tests comparing field toxicity generated from Chironomidae abundance and various hazard metric toxicity predictions are shown for Silt (Fig. 3) and LSand (Fig. 4) habitats with regression statistics included in Table 2. In addition, the analogous tests for IBI values are shown for Silt (Fig. 5) and LSand (Fig. 6) with regression statistics shown in Table 3.

For the Silt habitat, the HZD metric generated a higher coefficient of determination ($R^2 = 0.29$; $p = 0.001$) compared to PEC-Q methods ($R^2 = 0.03$ to 0.18 ; $p = 0.015$ to 0.308). In addition, the slope of the regression line ($m = 0.51$) more closely fit the 1:1 expectation indicating that it is had lower bias compared to the three PEC-Q methods. For the PEC-Q approaches, the PEC-Q_{avg} was generally too conservative. Several sites generated low toxicity predictions but also contained very low abundance or a complete absence of the benthic indicator. The PEC-Q_{sum} was overly sensitive with several sites predicted to have 100% toxicity while field abundance of chironomids were relatively high. For the IBI values, the PEC-Q_{avg} and PEC-Q_{met} resulted in significant negative correlations of $R^2 = 0.316$ ($p = 0.000$) and $R^2 = 0.323$ ($p = 0.000$), respectively, with the HZD metric of $R^2 = 0.101$ significant at the $p < 0.1$ level ($p = 0.071$). Unlike the Chironomidae abundance, there is no expected 1:1 relationship between the SQG scores and IBI values, thus the proximity of the slope to a value of 1 provides no guidance to

Revised March 12, 2016

evaluating goodness of fit tests. Overall, both the Chironomidae abundance (field toxicity) and IBI approaches showed reasonable correlations with the HZD, PEC-Q_{avg} and PEC-Q_{met} SQG scoring metrics. Thus, within the Silt habitat type, SQG-based toxicity scores are successful at predicting biological impact using either an indicator species or multimetric community composition (IBI) approach. The HZD was the strongest predictor of Chironomidae abundance, while PEC-Q_{met} explained the greatest variation in IBI scores.

For LSand, the HZD metric explained a similar amount of the variation of the field toxicity data ($R^2 = 0.27$; $p = 0.000$) as the PEC-Q methods ($R^2 = 0.25$ to 0.27 ; $p = 0.000$ to 0.001). Again PEC-Q_{avg} and PEC-Q_{met} were found to be overly conservative and predicted low toxicity despite numerous sites with 100% field toxicity. The HZD approach yielded a goodness of fit slope of 0.65 and most closely approximated the 1:1 correlation between predicted and field toxicity values. However this metric also tended to under predict toxicity at many sites. The PEC-Q_{sum} yielded a goodness of fit test slope of 0.39, lower than HZD but this was largely due to higher toxicity prediction at a few of the low contamination sites. For the IBI values, the HZD, PEC-Q_{avg} and PEC-Q_{met} all resulted in significant correlations of $R^2 = 0.391$ ($p = 0.000$), $R^2 = 0.407$ ($p = 0.000$) and $R^2 = 0.407$ ($p = 0.000$), respectively, while PEC-Q_{sum} had the lowest R² value of 0.130. As with the Silt habitat, the LSand Chironomidae toxicity and IBI approaches showed reasonable correlations with the HZD, PEC-Q_{avg} and PEC-Q_{met} (as well as PEC-Q_{sum}) SQG scoring metrics. In this case, HZD provided the strongest predictive capability for Chironomidae abundance (followed closely by PEC-Q_{sum}) while HZD, PEC-Q_{avg} and PEC-Q_{met} scoring methods explained similar amounts of variation in IBI score values.

The differences in the ability of the various hazard metrics to estimate abundance of Chironomidae between habitats may be suggestive of differences in chemical bioavailability.

Revised March 12, 2016

Hence, Chironomidae are apparently more sensitive with respect to their abundance in LSand habitats as compared to Silt habitats for the same degree of sediment contamination. This is consistent with higher organic carbon content in Silt contributing to a higher degree of chemical sequestering and lower overall contaminant bioavailability. However, these differences may also be a product of an oversimplification of the habitat classification method. While the selection of variables included in the habitat model were demonstrated to be relevant habitat predictors within the literature, benthic invertebrates within the study system may track habitats at a finer scale than what classification would model would suggest and errors in true habitat classification would contribute to additional variation in the goodness of fit test. This could be one reason for the failure to find any associations between priority contaminant concentrations and indicator abundances in the Grav and HSand habitats.

Differences between toxicity score predictions across the hazard metrics as contrasted in this study are due to the differences in their computation. Averaging hazard quotients across chemicals ($PEC-Q_{avg}$) results in conservative toxicity estimates, particularly when several chemicals are found in sediments at low concentrations relative to their $PEC-Q$ values. Thus, averaging causes 'dilution' of the overall score and consistently produces the lowest toxicity prediction. For the Chironomidae indicator in Silt and LSand, the $PEC-Q_{avg}$ was the poorest predictor of indicator abundance. Alternatively, $PEC-Q_{sum}$ can result in overly sensitive predictions of toxicity when several chemicals are present at low concentrations. In this case, the likelihood of exceeding an effect criterion, even when no field toxicity occurs, increases as the number of chemicals added to the metric increases. $PEC-Q_{sum}$ overestimated Chironomidae toxicity in Silt habitats but provided the second best prediction of indicator abundances in the LSand habitat where indicator sensitivity to contamination was apparently higher. Both the $PEC-$

Revised March 12, 2016

Q_{avg} and PEC- Q_{sum} could be improved by adopting the TEC-censoring algorithm used within the HZD methodology. Censoring chemicals that are present in sediments at concentrations less than the TEC value would increase the overall score for PEC- Q_{avg} and decrease that of PEC- Q_{sum} . However, the dilution artifact introduced by PEC- Q_{avg} can still present problems especially under the extreme case where only a single contaminant is present at high enough concentrations to elicit strong toxicological response. The PEC- Q_{met} provides an intermediate toxicity prediction given that metals are treated as a single contaminant while variations in organic chemical scores are given more weight in the algorithm. The PEC- Q_{met} is expected to be more appropriate when organic contaminants contribute to a larger proportion of sediment toxicity. It was the second best predictor of chironomid abundances in the Silt habitat and 3rd best predictor in LSand habitats whereas this metric was the strongest predictor of IBI in Silt but poorest predictor in LSand. PEC- Q_{met} could also benefit from adopting the TEC censoring method.

For the HZD method, its use a sigmoidal toxicity curve contributes to further differences in toxicity predictions compared to each of PEC- Q_x approaches. Figure S6 contrasts % toxicity predicted for a single contaminant (PCBs) over a theoretical range of sediment concentrations using HZD (Equation 6 and the TEC censoring algorithm) and PEC- Q_{PCB} . As demonstrated by the figure, HZD generates a lower toxicity estimate compared to PEC- Q_{PCB} when sediment concentrations of PCBs are between the TEC and PEC value and a higher estimate of toxicity for sediment concentrations exceeding PEC up to the 93% toxicity estimate where both toxicity curves intersect and eventually become maximized at 100% toxicity. When applied on a multi-chemical basis, HZD will therefore tend to produce lower toxicity estimates when several chemicals are less than PEC and higher toxicity estimates when multiple chemicals exceed their respective PEC. The HZD score provided the best overall predictor of chironomid abundances

Revised March 12, 2016

for the Silt and second best predictor for LSand, while for the IBI it was the poorest predictor for Silt and 4rd best predictor for LSand. Relative to PEC-Q_x approaches, the HZD suffers from its greater complexity in its computation.

Finally, each hazard metric approach can be potentially improved by calibration. In the present study, the PEC concentration was arbitrarily set to a 50% toxicity value. However, the value of toxicity associated with the PEC concentration can be set to other toxicity values and potentially optimized to establish a best fit to a calibration data set. Preliminary trials were performed using the HZD score by adjustment of PEC toxicity equivalents in the range of 25 to 75%. However, these adjustments were not found to make substantial improvements in goodness to fit test outcomes for chironomid abundances in this system. Furthermore, given the differences in sensitivity of chironomids to sediment contamination between habitats it is clear than no single optimized value is likely to work across all environments nor would this be applicable to other types of benthic indicator species.

Table 4 presents estimates of SQG-based toxicity score values for each of the computed metrics that correspond with the 50% field toxicity estimate for Chironomidae abundance in Silt and Sand habitats of the Detroit River based on goodness of fit tests. Thus, for Silt type habitats, stations having a computed HZD score greater than 54.1% might be considered potentially toxic with respect to chironomid abundances, whereas in LSand environments, a lower HZD score of 32.3% or greater might be considered potentially toxic. Table 4 also presents SQG-based toxicity metric scores that correspond to the median IBI scores achieved in these two habitat types. In this case, a HZD score of 65.2% and 50.6% will generate a median IBI scored for each respective habitat type. Such information could prove useful for setting site and habitat specific criteria or remediation goals using a multi-chemical metric approach. Thus, in Silt, HZD scores

Revised March 12, 2016

exceeding 50% are more likely to be considered toxic both from an indicator abundance perspective as well as have a higher likelihood of achieving lower IBI scores. In LSand habitats HAZ scores exceeding 30% may be useful for designating the potential for biological impacts. Table 4 also summarizes the score values for 50% Chironomid toxicity and scores that achieve the median IBI in Silt and LSand for PEC-Q_{avg}, PEC-Q_{met} and PEC-Q_{sum} metrics. In the case of PEC-Q_{avg} the degraded condition values for indicator abundance and IBI range from 11.9 – 14.4 % for Silt and 7.7 - 11.6% for LSand. Converting these values back to standard hazard quotient scales (achieved by dividing the %toxicity value by 50) yields PEQ-Q_{avg} on an absolute scale in the range of 0.24-0.29 and 0.15-0.23. Notably, these protective values are much lower than the value of 1 commonly used in hazard quotient assessment methods.

Conclusion

SQGs such as those implemented by MDEQ are useful for providing toxicity benchmarks on which to evaluate sediment contamination and potential of sediments to contribute to degraded benthos. However, these guidelines cannot be used appropriately without considering the interactive effects of mixtures of chemicals present in the field and different habitats present within the system. Various multi-contaminant toxicity score approaches have been developed. However, different scoring methods generate different estimates of sediment toxicity and each computational approach contains within it assumptions that can contribute to biases in the prediction. This study demonstrated the applicability of some common multi-chemical multivariate scoring approaches for predicting benthic invertebrate abundance and indices of biological integrity (IBI) under field conditions. It further describes a new method (HZD) that adopts a multi-point calibration through incorporation of information provided by PEC and TEC

Revised March 12, 2016

benchmark values. The HZD score and individual PEC-Q approaches were capable of predicting Chironomidae abundances and IBI score values in two types of Detroit River habitats with varying degrees of success. The HZD approach provided the best estimate of chironomid abundance in silt habitat, while PEC-Q_{sum} produced the best estimate of chironomid abundance in low flow sand habitats followed closely by the HZD score as the second best metric. As for the multivariate approaches, the multimetric IBI approach showed correlations in the Silt and LSand habitats for each of the SQG scoring metrics. However, none of the metrics were able to predict abundances of other organism groups and habitat types indicating some limitations of the toxicity interpretation of SQGs across habitat types and species. This implies that habitat and taxonomic specific SQGs are likely warranted. Overall, the universality of SQGs is generally limited by: (1) the types of organisms present in the local environment and their tolerance to contaminant and habitat quality; (2) differences in contaminant availability based on site physical and chemical characteristics; and (3) unknown toxicological interactions of multiple contaminants at a site. Overall, the SQG LOE screening can be useful as part of a full WOE assessment of sites to identify impacts of contamination on the sediment benthos. The method is made more powerful when habitat characteristics are considered within the SQG LOE assessment process and this has the added advantage of adopting consistency in approach with other LOEs that exclusively consider biological community composition alterations relative to habitat matched reference locations.

ACKNOWLEDGMENTS

Funding for this project was provided by Great Lakes Sustainability Fund, Environment Canada and Canada-Ontario Agreement Funds, Ontario Ministry of Environment and Climate Change to

Revised March 12, 2016

G.D. Haffner and K.G. Drouillard. The authors would like to thank Sarah Wood who completed the benthic invertebrate taxonomy, Maciek Tomczak who was involved in planning and implementation of the Detroit River sediment sampling survey, Rodica Lazar who completed chemical analysis and Todd Leadley for implementing components of the field survey. We would also like to acknowledge the useful contributions to improving this manuscript made by two anonymous reviewers.

REFERENCES

- Bay SM, Weisberg SB. 2010. Framework for interpreting sediment quality triad data. *Integr Environ Assess Manag* 8(4):589-596.
- Bay SM, Ritter KJ, Vidal-Dorsch DE, Field LJ. 2012. Comparison of national and regional sediment quality guidelines for classifying sediment toxicity in California. *Integr Environ Assess Manag* 8(4):597-609.
- Besser JM, Giesy JP, Kubitz JA, Verbrugge DA, Coon TG, Braselton WE. 1996. Assessment of sediment quality in dredged and undredged areas of the Trenton Channel of the Detroit River, Michigan USA, using the sediment quality triad. *J Great Lakes Res* 22(3):683-696.
- Chapman PM. 1996. Presentation and interpretation of Sediment Quality Triad data. *Ecotoxicology* 5(5):327-339.
- Chapman PM, Anderson J. 2005. A decision-making framework for sediment contamination. *Integr Environ Assess Manag* 1(3):163-173.
- Drouillard KG, Tomczak M, Reitsma S, Haffner GD. 2006. A river-wide survey of polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), and selected

Revised March 12, 2016

- organochlorine pesticide residues in sediments of the Detroit River - 1999. *J Great Lakes Res* 32(2):209-226.
- Farara DG, Burt AJ. 1993. Environmental assessment of Detroit River sediments and benthic macroinvertebrate communities, 1991. Report prepared for the Ontario Ministry of the Environment by Beak Consultants Limited, p. 203, Brampton, Ontario.
- Ingersoll CG, MacDonald DD, Wang N, Crane JL, Field LJ, Haverland PS, Kemble NE, Lindskoog RA, Severn C, Smorong DE. 2001. Predictions of sediment toxicity using consensus-based freshwater sediment quality guidelines. *Arch Environ Contam Toxicol* 41(1):8-21.
- Long ER, Ingersoll CG, MacDonald DD. 2006. Calculation and uses of mean sediment quality guideline quotients: A critical review. *Environ Sci Technol* 40(6):1726-1736.
- MacDonald DD, Ingersoll CG, Berger TA. 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. *Arch Environ Contam Toxicol* 39(1):20-31.
- Parsons M, Norris RH. 1996. The effect of habitat-specific sampling on biological assessment of water quality using a predictive model. *Freshwater Biol* 36(2):419-434.
- Rae JG. 1985. A multivariate study of resource partitioning in soft bottom lotic Chironomidae. *Hydrobiologia* 126(3):275-285.
- Reitsma S, Drouillard KG, Haffner D. 2003. Simulation of sediment dynamics in Detroit River caused by wind-generated water level changes in Lake Erie and implications to PCB contamination. In *Evaluating Ecosystem Results of PCB Control Measures Within the*

Revised March 12, 2016

- Detroit River-Western Basin* Heidtke, T., Hartig, J.H. and Yu, B. (eds), Great Lakes National Program Office, U.S. Environmental Protection Agency, Chicago, IL.
- Reynoldson TB, Bailey RC, Day KE, Norris RH. 1995. Biological guidelines for freshwater sediment based on Benthic Assessment of Sediment (the BEAST) using a multivariate approach for predicting biological state. *Austral J Ecol* 20(1):198-219.
- Reynoldson TB, Norris RH, Resh VH, Day KE, Rosenberg DM. 1997. The reference condition: A comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *J N Am Benthol Soc* 16(4):833-852.
- Reynoldson TB, Rosenberg DM, Resh VH. 2001. Comparison of models predicting invertebrate assemblages for biomonitoring in the Fraser River catchment, British Columbia. *Can J Fish Aquat Sci* 58(7):1395-1410.
- Ritter KJ, Bay SM, Smith RW, Vidal-Dorsch DE, Field LJ. 2011. Development and evaluation of sediment quality guidelines based on benthic macrofauna responses. *Integr Environ Assess Manag* 8(4):610-624.
- Szalinska E, Drouillard KG, Fryer B, Haffner GD. 2006. Distribution of heavy metals in sediments of the Detroit River. *J Great Lakes Res* 32(3):442-454.
- Wenning RJ, Ingersoll CG, Batley GR, Moore DW. 2004. Use of Sediment Quality Guidelines (SQGs) and Related Tools for the Assessment of Contaminated Sediments, SETAC Press (Society of Environmental Toxicology and Chemistry), Pensacola, FL.
- Wright JF. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Austral J Ecol* 20(1):181-197.

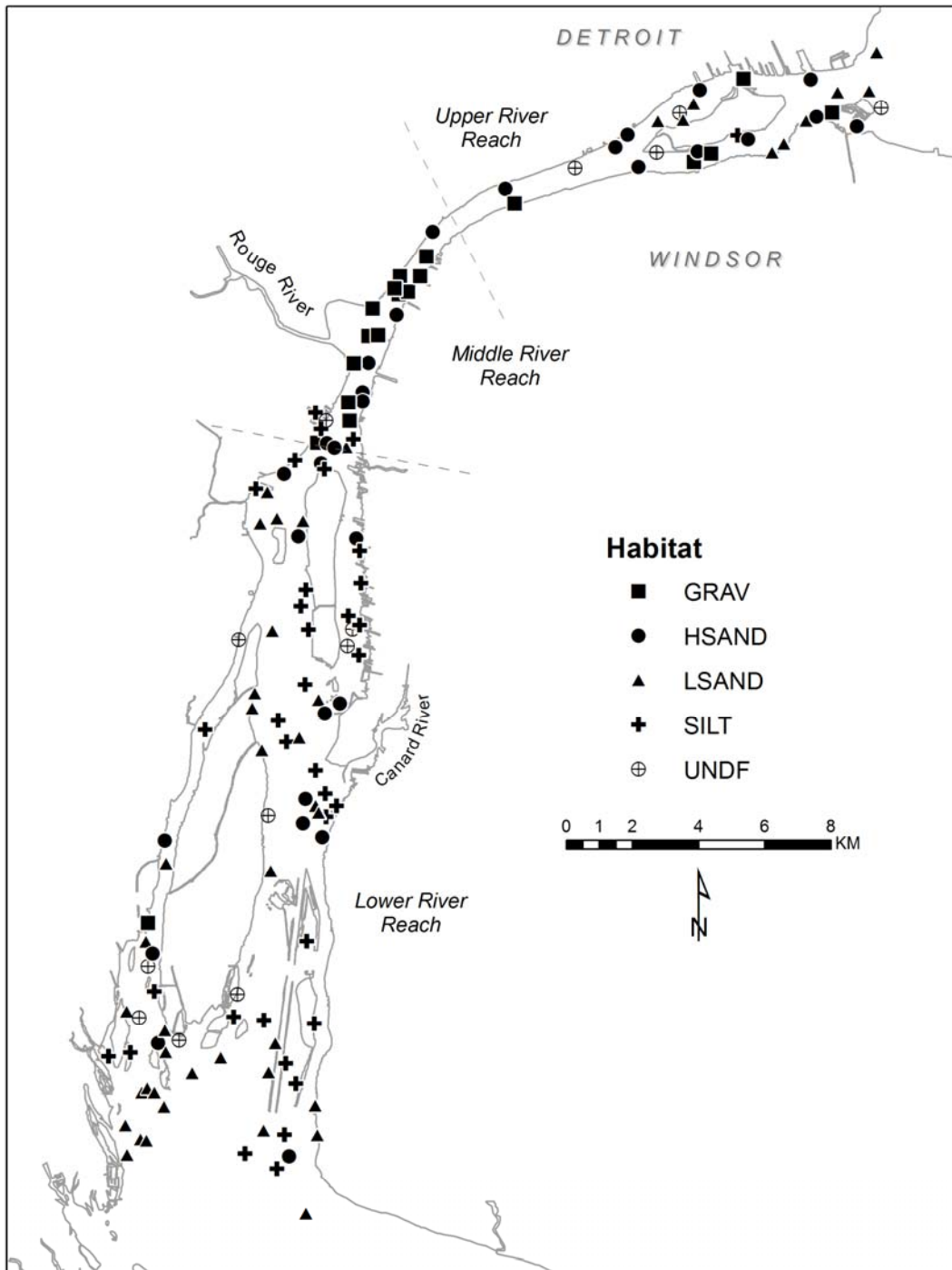


Figure 1: Sampling location habitats identified after discriminant function analysis (DFA) along the Detroit River.

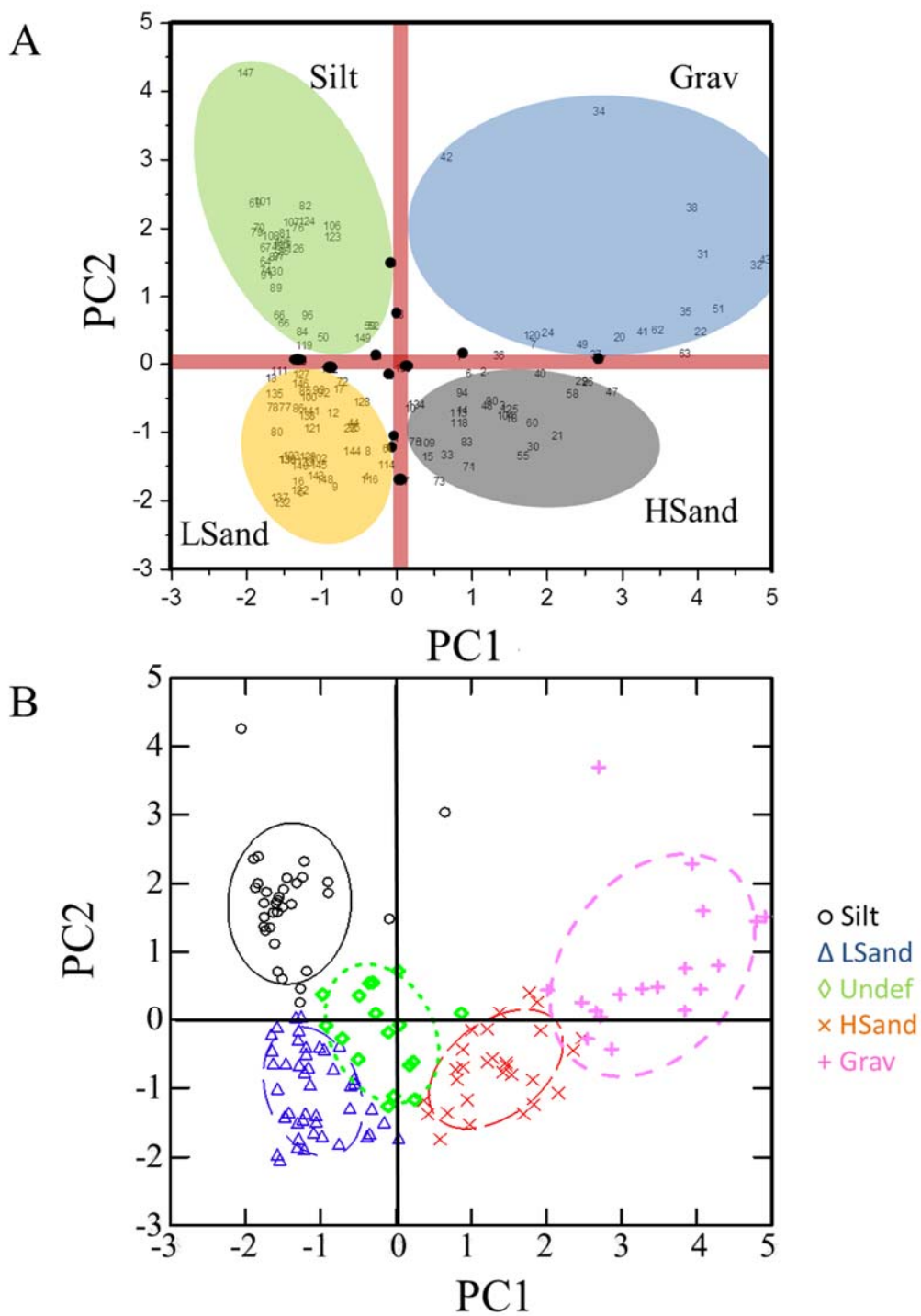


Figure 2: (A) Results of principle component analysis (PCA) with user-identified habitat ellipses used for initial habitat characterization prior to discriminant function analysis (DFA_ (numbers indicate sampling locations). Black circles denoted as undefined habitats; (B) Results of PCA

Revised March 12, 2016

with habitat variables identified using DFA. Ellipses represent 99% confidence intervals around a given habitat type.

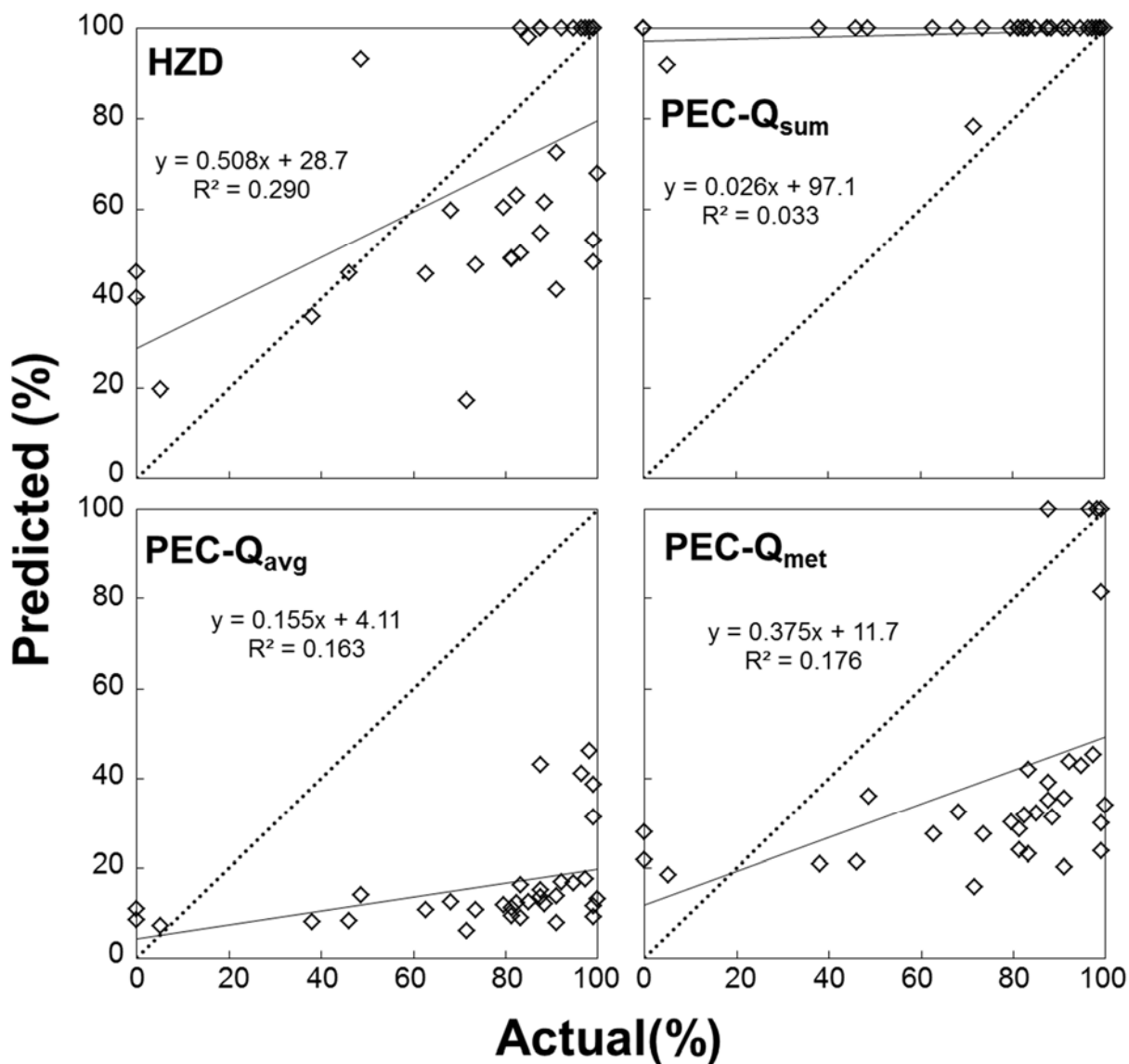


Figure 3: Chironomid toxicity versus SQG scores from four approaches (*HZD*, *PEC- Q_{avg}* , *PEC- Q_{met}* and *PEC- Q_{sum}*) for Silt habitat. Dashed line indicates a 1:1 correlation between actual measured and predicted toxicities.

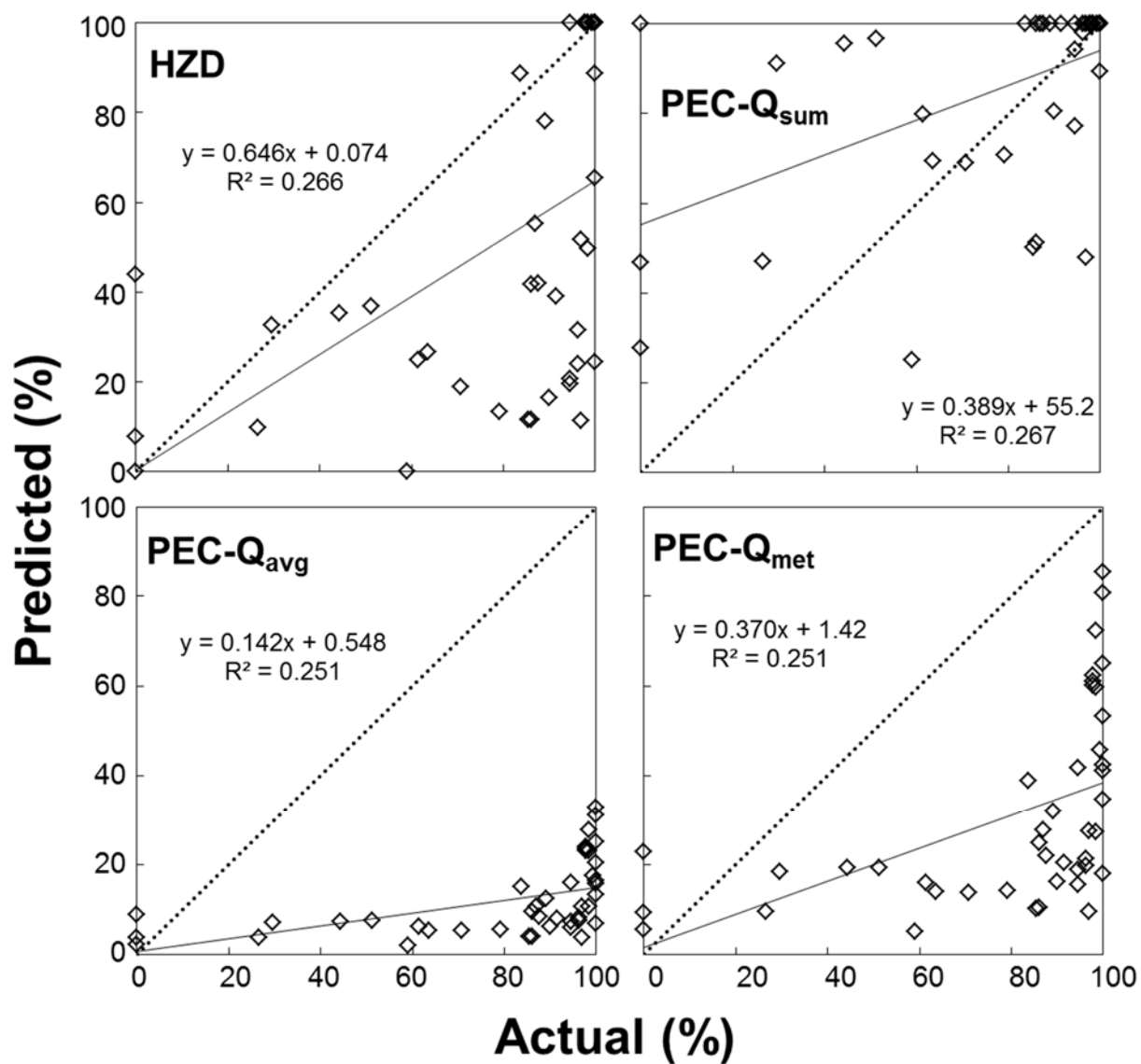


Figure 4: Chironomid toxicity versus SQG scores from four approaches (*HZD*, *PEC- Q_{avg}* , *PEC- Q_{met}* and *PEC- Q_{sum}*) for LSand habitat. Dashed line indicates a 1:1 correlation between actual measured and predicted toxicities.

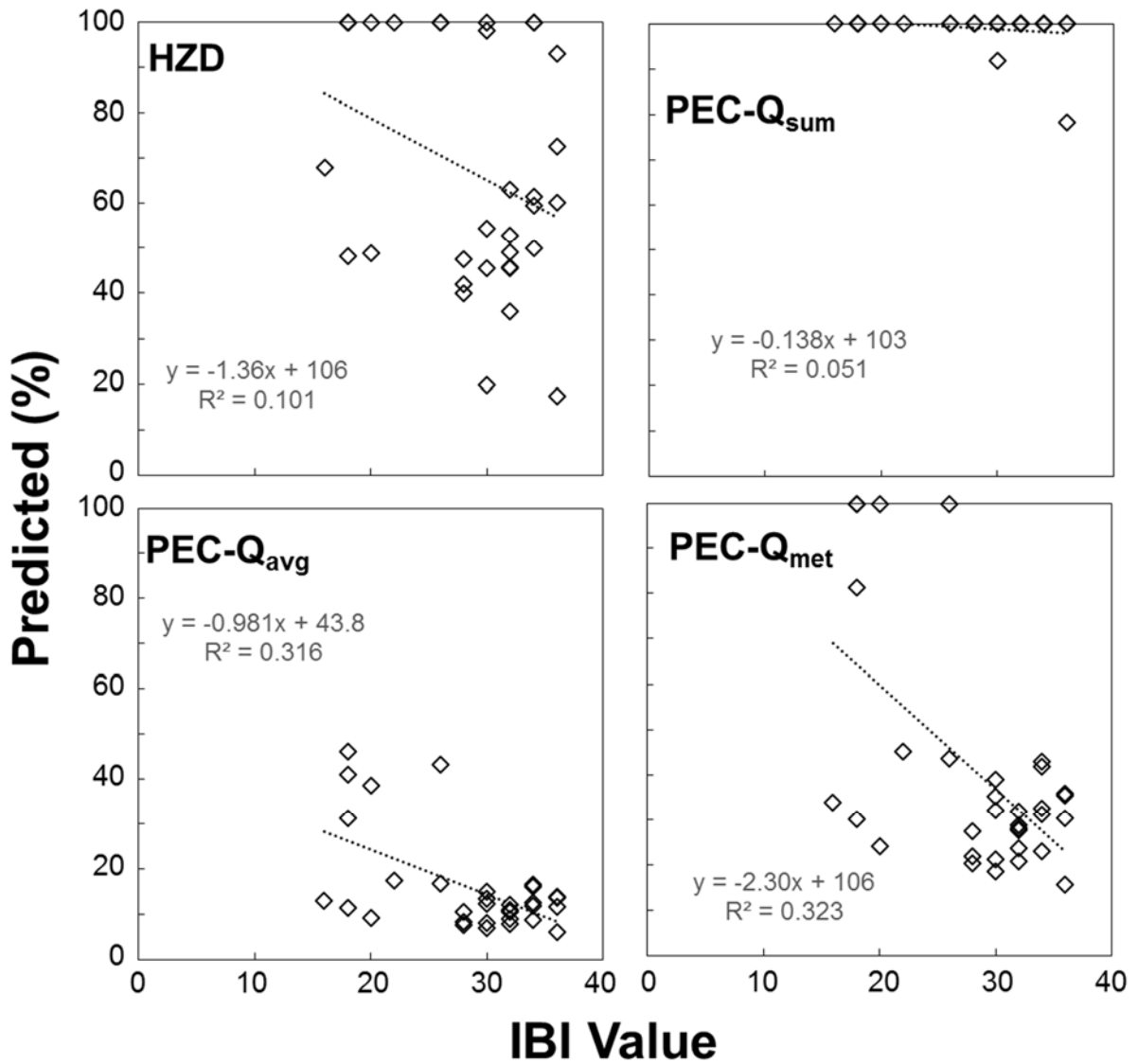


Figure 5: IBI values versus SQG scores from four approaches (*HZD*, *PEC- Q_{avg}* , *PEC- Q_{met}* and *PEC- Q_{sum}*) for Silt habitat.

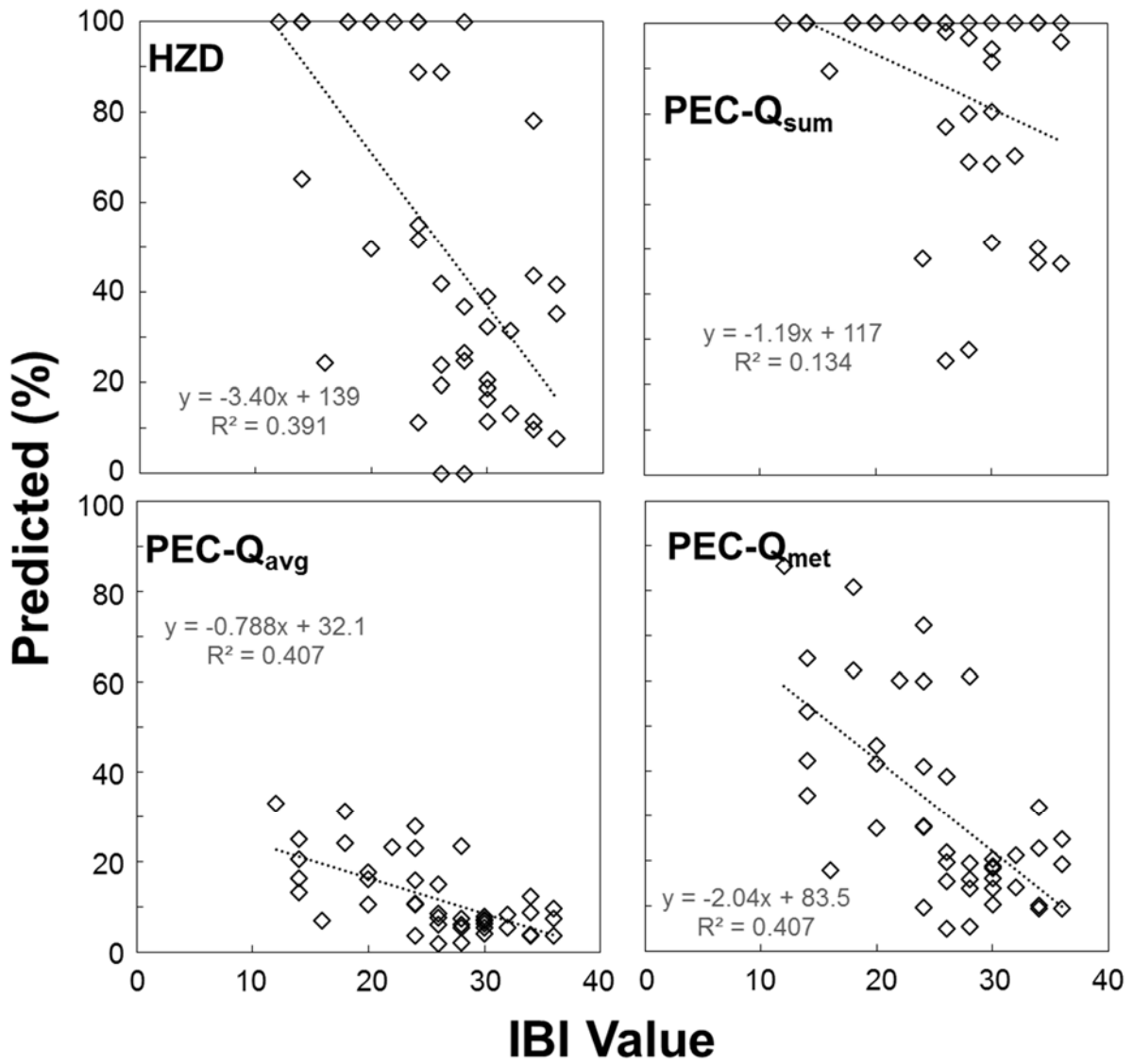
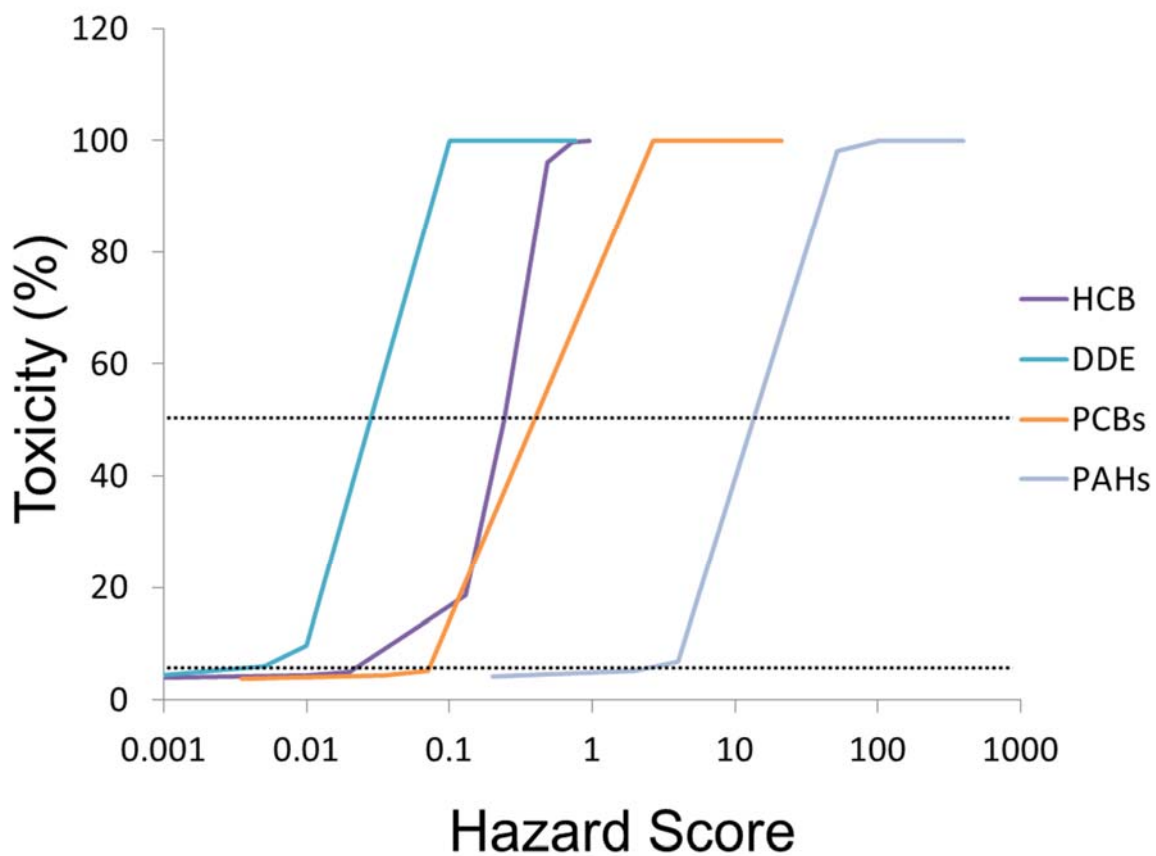
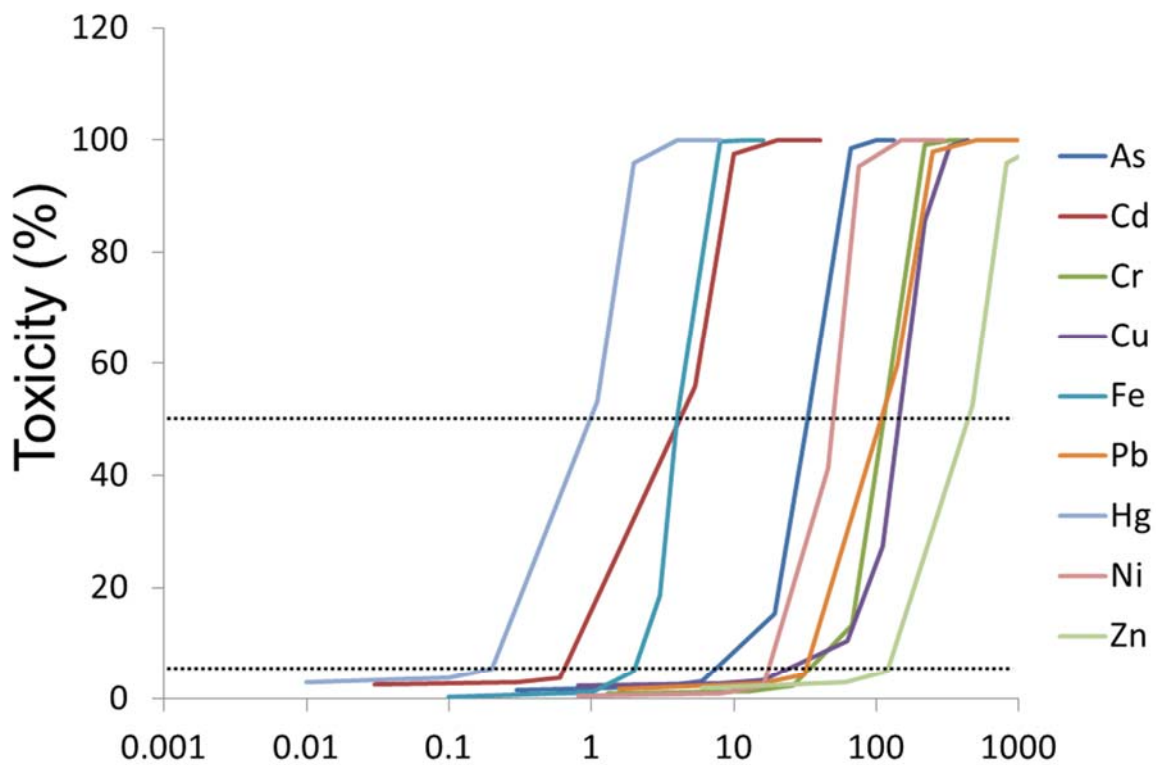


Figure 6: IBI values versus SQG scores from four approaches (*HZD*, *PEC- Q_{avg}* , *PEC- Q_{met}* and *PEC- Q_{sum}*) for LSand habitat.

Revised March 12, 2016

Supplemental Information

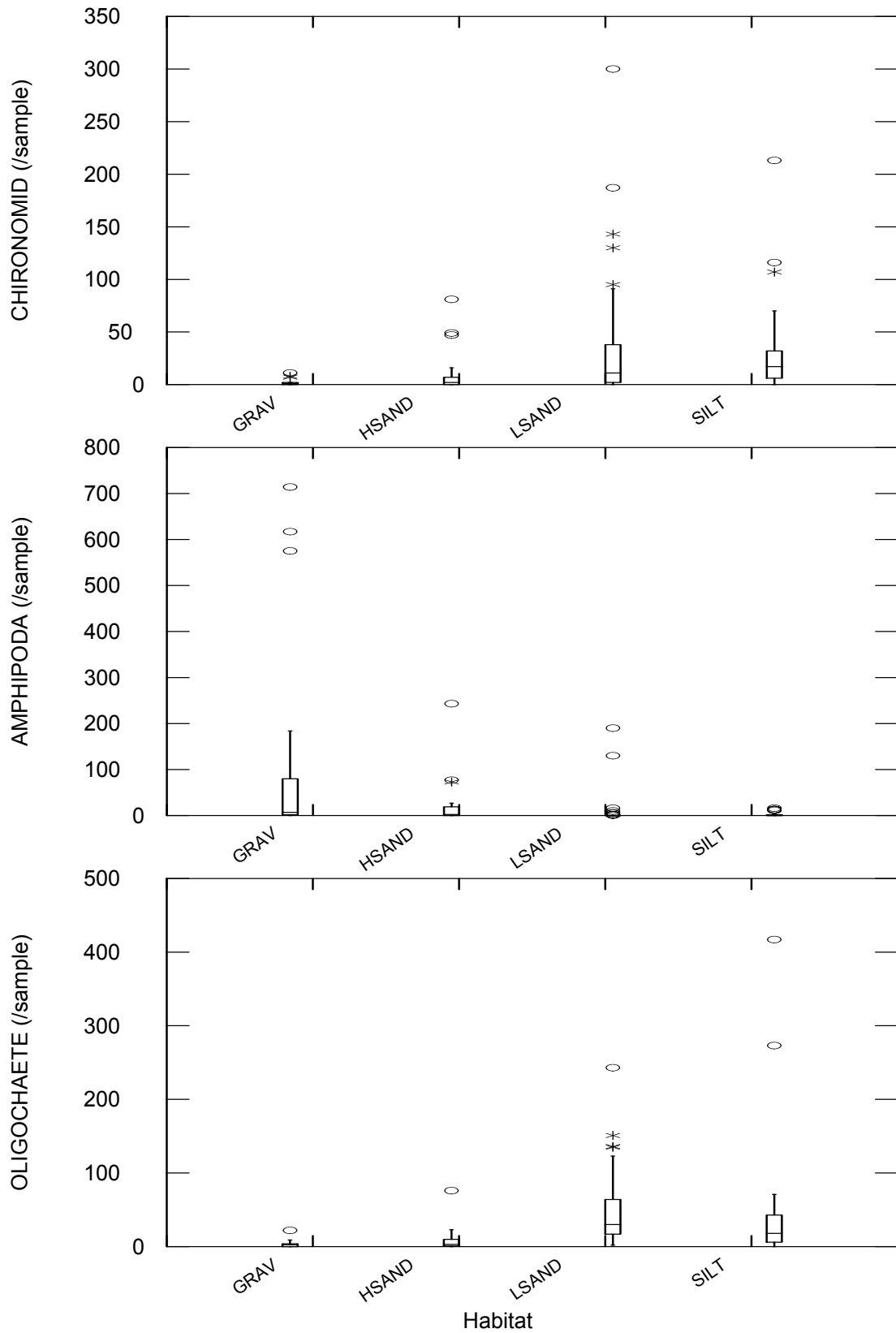
Revised March 12, 2016



Revised March 12, 2016

Figure S1: Dose-response curves created using the LEL and SEL concentrations from Table S1 for the Hazard Scoring approach.

Revised March 12, 2016



Revised March 12, 2016

Figure S2: Abundances of dominant species (Oligochaete, Amphipods and Chironomids) for each of four habitats (Grav, HSand, LSand and Silt).

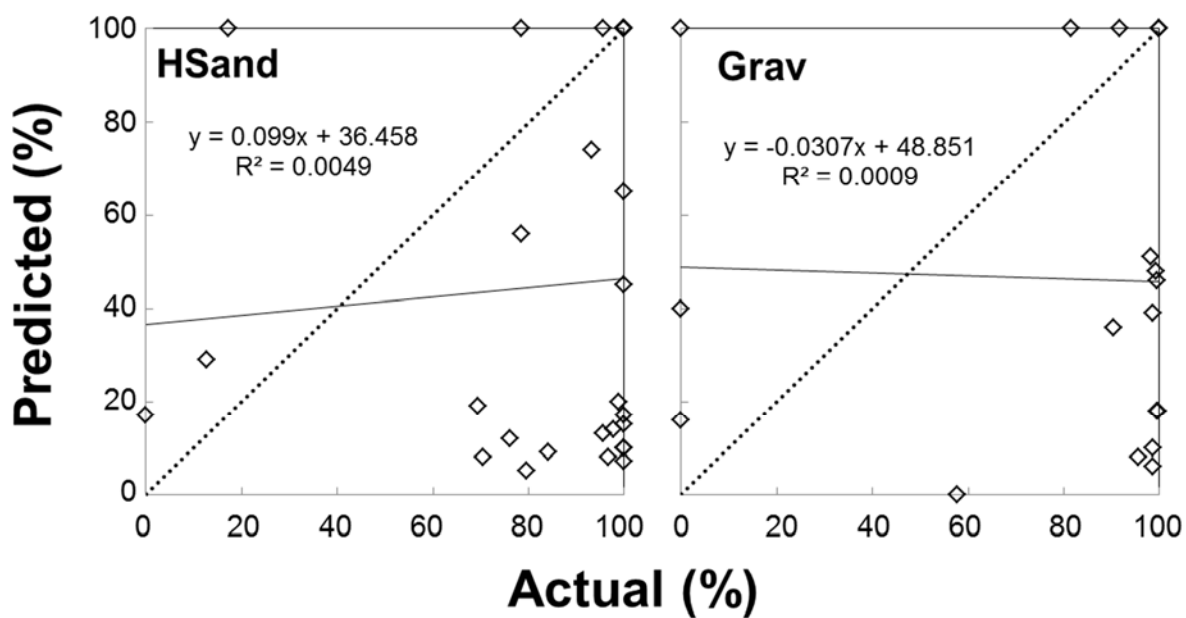


Figure S3: Amphipoda toxicity versus *HZD* scores for two habitats with low Chironomid abundances (HSand and Grav). Dotted line indicates a 1:1 correlation between actual measured and predicted toxicities.

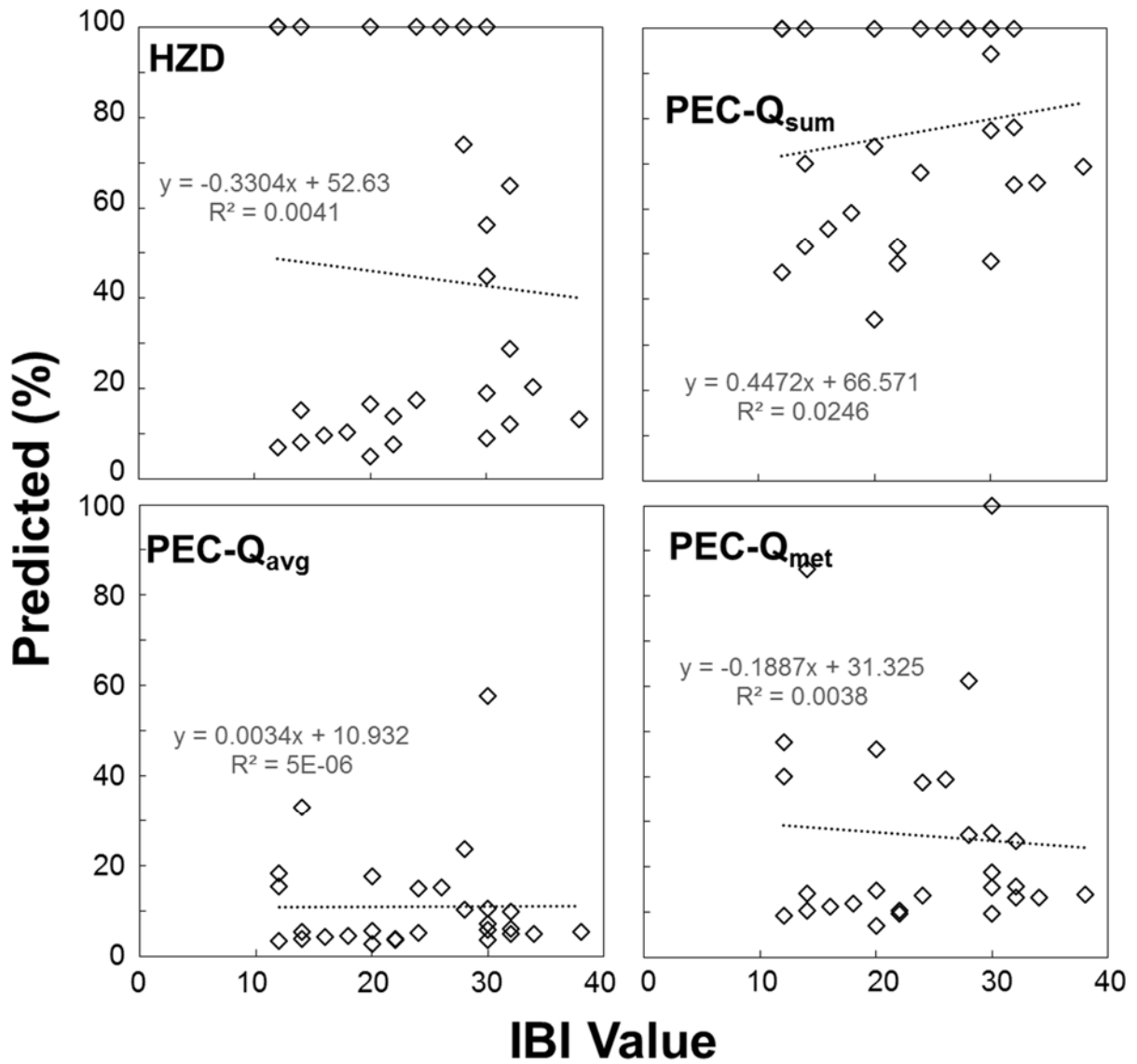
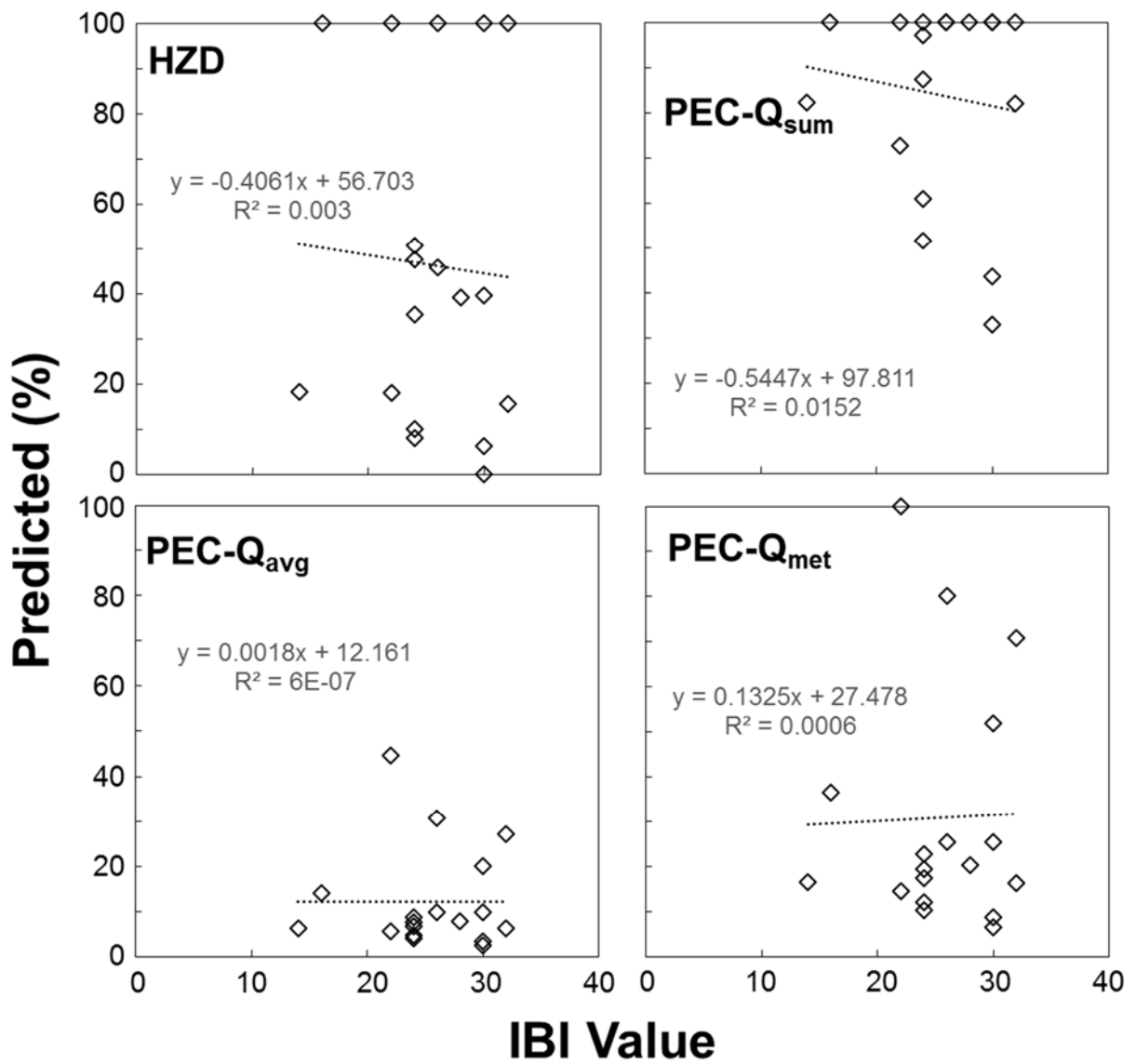


Figure S4: IBI values versus SQG scores from four approaches (HZD, PEC-Q_{avg}, PEC-Q_{met} and PEC-Q_{sum}) for the HSD habitat.



Revised March 12, 2016

Figure S5: IBI values versus SQG scores from four approaches (HZD, PEC-Q_{avg}, PEC-Q_{met} and PEC-Q_{sum}) for the Grav habitat.

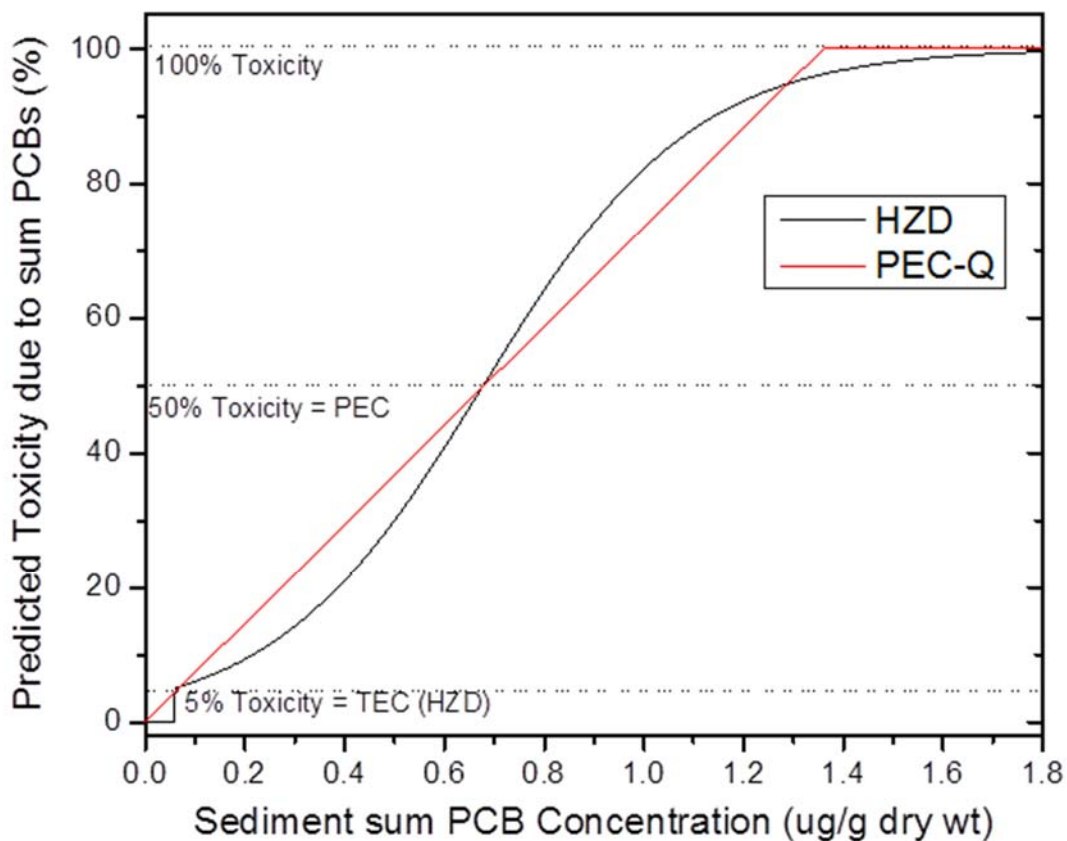


Figure S6: Example toxicity curves for sum PCBs versus sediment concentration. The Hazard Score (HZD) is 0% until the concentration reaches 5% toxicity at the TEC and conforms to a sigmoidal toxicity curve. The PEC-Q_x toxicity is linear from 0 to 100% and capped at 100% toxicity.

Table S1: SQG method chemical concentrations and fitted sigmoidal curve parameters.

	Chemical													
	parameter	As	Cd	Cr	Cu	Fe	Pb	Hg	Ni	Zn	HCB	DDE	Total PCBs	Total PAHs
MDEQ	TEC	9.79	0.99	43	32	2	36	0.18	23	121	2.00E-02	3.16E-03	0.06	1.61
	PEC	33	5	111	149	4	128	1.06	49	459	2.40E-01	3.13E-02	0.68	23
	fitted k	0.13	0.74	0.04	0.03	1.47	0.03	3.35	0.11	0.01	13.38	105	4.75	0.14
	fitted A	66	40	126	42	361	60	35	251	55	25	26	25	24
	Ec	1.5	2.5	0.8	2.3	0.3	1.7	2.8	0.4	1.8	3.9	3.6	3.8	4.1

Table S2: Results of PCA for various habitat parameters after initial analysis.

Parameter	Component Loadings	
	PC1 (50.1%)	PC2 (28.4%)
Depth (m)	-0.884	0.002
Velocity (m/s)	-0.900	-0.071
TOC (%)	0.092	0.779
Gravel (%)	-0.905	0.137
Sand (%)	0.528	-0.761
Silt (%)	0.557	0.703

Bold values indicate strongly loaded values ($p > 0.7$).

Table S3: Results of DFA for categorization of habitats.

Classification Matrix (Cases in row categories classified into columns)

Revised March 12, 2016

	GRAV	HSAND	LSAND	SILT	UNDF	%correct
GRAV	17	1	0	0	0	94
HSAND	0	27	0	0	1	96
LSAND	0	0	43	0	2	96
SILT	0	0	1	31	1	94
UNDF	0	1	0	0	16	94
Total	17	29	44	31	20	95

Table S4: Summary of species representing $\geq 0.5\%$ of total organism abundance and results of final PCA analysis.

	PC1	PC2	PC3
	(41%)	(17%)	(15%)
Oligochaeta	-0.058	0.030	-0.935
Nematoda*			
Chironomidae	0.027	-0.806	-0.259
Ceratopogonidae*			
Hexagenia	-0.099	-0.737	0.273
Caenis*			
Hydropsychidae*			
Other Trichoptera*			
Amphipoda	0.905	0.027	0.042
Dreissena*			

Revised March 12, 2016

Hydrozoa	0.823	0.004	0.029
Hirudinea*			
Turbellaria	0.771	0.054	0.044
Gastropoda	0.889	-0.134	-0.131

*denotes species not strongly correlated after initial PCA analysis and excluded from further analysis. Bolded PCA scores indicate strongly loaded variables on that particular component axis (< -0.7 and > 0.7)

Table S5: Values for IBI metrics used for scoring. (EPT = Ephemeroptera, Trichoptera, and Plecoptera).

Metric	Score	Values
Abundance	5	>143
	3	31-143
	1	<31
No. families	5	>6
	3	3-6
	1	<3
% EPT	5	>2.5
	3	0.01-2.5
	1	<0.01

Revised March 12, 2016

% Chironomids	5	>37
	3	0.1-37
	1	<0.1
No. EPT taxa	5	1
	1	0
% dominance	5	>51
	3	51-82
	1	<82
Shannon Wiener Index	5	>0.61
	3	0.29-0.61
	1	<0.29
Evenness	5	>0.74
	3	0.42-0.74
	1	<0.42