Author: Verschoor, Aletta Johanna

Title: The power of biotic ligand models : site-specific impact of metals on aquatic communities

Issue Date: 2013-12-04

Chapter 6

Discussion

Effects of metals on aquatic ecosystems

Metals are detected in every environmental compartment and are inherently persistent. They undergo various speciation changes within different compartments (Tessier et al., 1995) and tend to accumulate in biota (Phillips et al., 1994). Total metal concentrations measured in water often fail to predict ecological effects accurately, and therefore there is a need for models that account for bioavailability. Many studies have shown that metal toxicity depends on water type-specific characteristics, such as pH, DOC and hardness (De Schamphelaere et al., 2004a; Pagenkopf, 1983). These parameters explain differences in toxicity in surface waters with equal dissolved metal concentrations. Bioavailability is a useful concept to refine risk assessments, and to select and prioritize sites that are at risk.

One of the most promising models is the Biotic Ligand Model (BLM), so far developed for individual metals (see Chapter 1). Effects prediction using the BLM is dictated by water chemistry, including competition, metal availability and the effect concentration of the metal at the target site of toxic action. The idea behind BLMs is that a certain critical concentration should be reached at the epithelial binding site to trigger effects (Pagenkopf, 1983; Paquin et al., 2002a). The mode of action of metal binding on the biotic ligand in fish has been identified as disturbance of ionoregulation, followed by decreased Ca levels in the blood (hypocalcaemia), which is linearly related to initial effects and eventually results in mortality of the fish (Paquin et al., 2002b).

The BLM concept is applicable to a wide range of organisms, i.e. plants (Lock et al., 2007), algae (Deleebeeck et al., 2009a), crustacean (Bossuyt et al., 2004), snails and rotifers (De Schamphelaere et al., 2010) and fish (De Schamphelaere et al., 2004b) and can be coupled to species sensitivity distributions (Posthuma et al., 2002). Taking bioavailability into account in this way, improves the ability to generate site-specific water quality criteria (Niyogi et al., 2004; Vijver et al., 2008).

In the Netherlands, the scope for implementing the BLM of Cu was being explored and the model has been subjected to a sensitivity and uncertainty analysis (Vijver et al., 2008). PNECs were calculated for 6 water types. The PNEC of 2.4 µg/L, derived for the most vulnerable water type, was adopted[1] as a generic quality standard, to be implemented in legislation in 2015. The wide-spread use of BLMs by environmental regulators however, was hampered by the complexity of the BLM procedure and the large number of required input parameters. Another limitation of BLMs, affecting their usefulness as a field impact prediction tool, was that they are derived in the laboratory and for individual metals - whereas metals occur in the field by definition in mixtures. Moreover, in addition to abiotic factors, many other parameters related to the habitat of species vary. These factors include the food web structure of communities, their productivity, exposure history, life history and disturbance regime (Clements et al., 2012).

The aims of this PhD-thesis were:

1. to verify and optimize the ability of biotic ligand models to predict effects under realistic field conditions
2. to facilitate implementation of site-specific risk assessment methodology for several metals, based on mechanistic descriptions of biotic ligand models

The aims are specified in four research questions, which are answered in separate chapters of this thesis.

1. Is there a significant relationship between the calculated biotic ligand binding of metals and the measured bioaccumulation in aquatic species in the field?
2. How accurately do single metal BLMs, extended with a mixture model, predict toxicity of metals in a field setting?

[1] Adopted by board of directors of the Ministry of Infrastructure and the Environment (DIRBOWA, November 2009) to be implemented in “Regeling Monitoring KRW” in 2015.

3. What are vulnerable conditions and time periods for metal exposed ecosystems, based on changes in water chemistry and calculated effects on metal bioavailability?
4. Is it possible to derive a simplified function, based on a limited number of monitoring parameters, to facilitate widespread, practical use and implementation of BLMs with acceptable predictive capacity?

This PhD assists in incorporation of BLMs in operational risk assessment. The results are made operational by 1) the development of calculation tools, 2) the assessment of uncertainties of risks related to natural variations in water chemistry and 3) optimization of BLM for field predictions.

This PhD-thesis primarily focused on Cu, Ni and Zn, because of their widespread occurrences, their legal status as (priority) pollutants in monitoring programs, and the presence of well-documented BLMs and toxicity databases.

Prior to an integrated discussion, answers to the research questions are highlighted below.

Answers to the research questions

Research question 1 (Chapter 2): Elevated concentrations of metals had a negative impact on growth of two different crustacean species, commonly present in natural surface waters. A reduced body weight was related to bioaccumulation of Cd, Co and Mn in *D. magna* and *G. roeseli.* Bioaccumulation was related to the occupancy of the biotic ligand.

Research question 2 (Chapter 3): BLMs were suitable to rank sites with respect to effects of metals on population growth of *D. magna.* Effects on population growth could be attributed to Cd and Zn in the *in situ* experiment and to Co and Ni in the laboratory experiment. The sensitivity of *D. magna* under these multimetal field exposure conditions was approximately 20 – 30 times higher, than in original BLMs.

Research question 3 (Chapter 4): Monitoring data were organized to derive seasonal patterns over a 3.5 years period. Mean seasonal variations of estimated metal-induced risks upto a factor 2, were caused by variations in the concentrations of dissolved metals and other water chemistry parameters. Highest risk were predicted in February, lowest risks in September/October, whereas May resembled the annual average risk. Knowledge of the seasonal variations enables the reduction of sampling frequency, while covering the minimum and maximum values of HC5s of metals.

Research question 4 (Chapter 5). Simplification of BLMs by linear regression showed that a high level of accuracy of predicted HC5 can be maintained, while the number of BLM-parameters is reduced. Only a limited number of input parameters was required: DOC, eventually extended with pH and Ca. Adding more water chemistry parameters obviously gave more accurate predictions, but did not improve HC5 prediction significantly.

In the following paragraphs three definitions for validity, variation and uncertainty are employed, which can be easily confused. Their definitions are:

Calibration factor: A factor describing the deviation between measurements and model outcomes.

Uncertainty factor: A factor describing the variation in a dataset

Safety factor: A political value, applied to cover other unquantified uncertainties.

In line with the first aim of the PhD project "to verify and optimize the ability of biotic ligand models to predict effects under realistic field conditions", the validity of BLMs for field predictions will be discussed. In line with the 2nd aim "to facilitate implementation of site-specific risk assessment methodology…" recommendations for data gaps in water chemistry databases, optimal sampling frequencies and safety factors in site specific risk assessment are presented.

Validity of biotic ligand model for field predictions

This thesis provides evidence from *in situ* experiments on 12 field locations and parallel experiments with the same water samples in the laboratory (Chapter 3). This type of experiments was not done before, and added new information with respect to the validity of BLMs to predict field effects. This paragraph discusses: 1) the options for BLM validation, 2) validation of chemical speciation modeling; 3) validation of competitive binding model to the biotic ligand, and 4) extrapolation of BLMs to other species (read across). Finally, conclusions and recommendations for research and policy are presented.

Options for BLM validation

Evaluation of the validity of BLMs was extensively discussed internationally, and was laid down in European risk assessment reports of Cu, Ni and Zn. For most of the natural surface water samples (De Schamphelaere et al., 2002a; De Schamphelaere et al., 2003; De Schamphelaere et al., 2004c; Schwartz et al., 2007) or organisms sampled in natural surface waters (Bossuyt et al., 2004; Muyssen et al., 2005), metal effects could be predicted within a factor 2 accuracy. The applicability of Daphnia BLMs for more sensitive species was also confirmed (De Schamphelaere et al., 2006; De Schamphelaere et al., 2010; Schlekat et al., 2010). The Scientific Committee on Health and Environmental Risks concluded that: “The use of BLMs […] provides sufficient protection from the potential effects of metals“ (SCHER, 2010).

Validation is defined as “determining whether a model predicts the effects well “. Validation is often accompanied by calibration of model parameters in order to achieve agreement between predicted values (model output) and measured data.

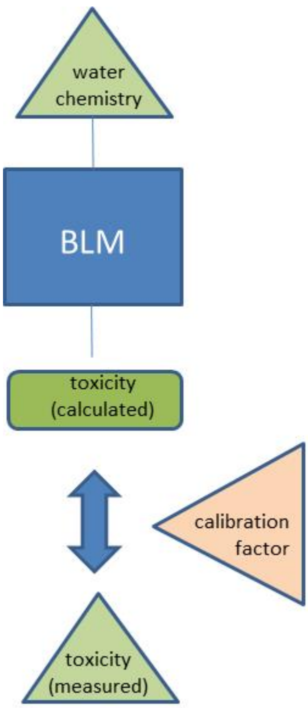


Figure 6.1 Scheme for functional validation. A step in this approach is the calculation of a calibration factor, to maximize the agreement between measured and modeled toxicity. Default BLM and speciation models are applied and their parameters are not adjusted.

Two types of validation can be distinguished: functional and conceptual validation. Functional validation involves a direct comparison of model output (EC_{50} or NOEC) with observations, without going into the details of underlying processes (See Figure 6.1). A functional validation provides insight in of the relevance and robustness of the model under more realistic conditions. The conditions in a validation study are meant to be more realistic than those applied for model derivation; i.e. natural surface water samples instead of synthetic media, or species sampled from natural habitats instead of species cultured in the laboratory, or even *in situ* exposure instead of lab-exposure.

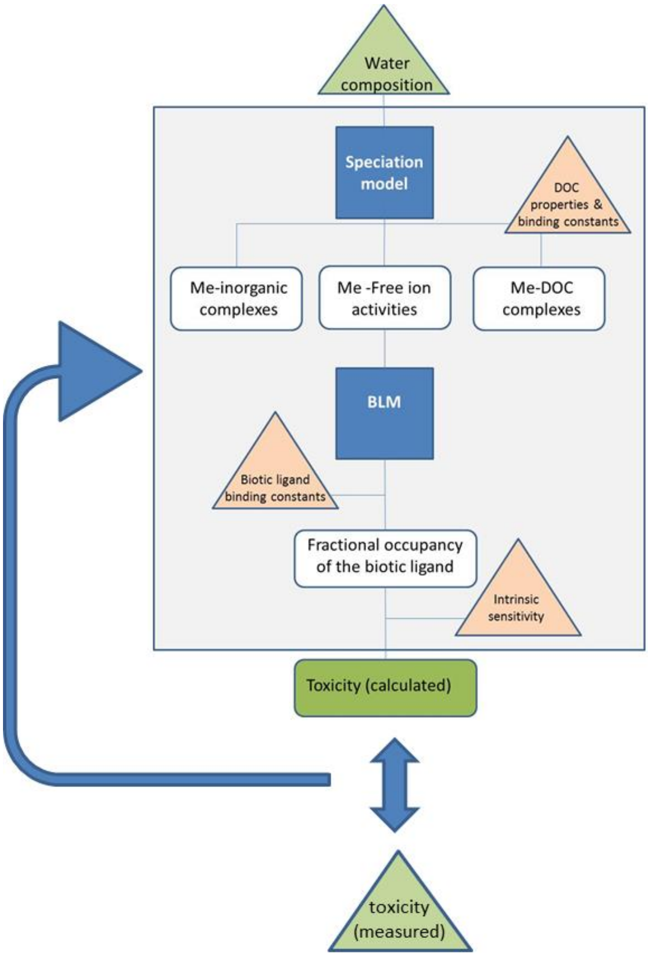


Figure 6.2 Scheme of conceptual calibration of BLMs. Models are visualized by blue squares, model input is visualized by triangles, model output by rectangles. Intermediate results are expressed in white rectangles. Orange triangles are model parameters, that are subject to calibration in validation studies. In a conceptual validation, intermediate results, for instance the free metal ion activity are compared with measurements in order to improve description of the speciation step.

In Chapter 3 it was observed that processes which are not included in the BLM cause a mismatch between model predictions and field observations. This mismatch was captured in a calibration factor, to account directly for deviations between model predictions and field observations. A functional validation was applied, and calibration factors of 20 for Co and Ni and 30 for Cd and Zn were calculated.

Conceptual validation involves an evaluation of the underlying processes in the model, for example the validity of the metal complexation binding to DOC or inorganic ligands, or the competition between cations (H, Ca, Na and Mg) and metals. Laboratory experiments are usually employed to ensure that the process of interest is not affected by other processes. Conceptual validation often goes along with calibration of model parameters, to improve the description of particular processes. In Figure 6.2 the modeling steps and calibration options for a conceptual approach are shown. During conceptual validation, specific speciation or BLM parameters are adjusted to maximize the agreement between measured and calculated toxicity, as was done for species sensitivity ($f_{BL,50\%}$) in Chapter 3. Potential causes for a deviation between predicted and observed toxicity endpoints are summarized in the textbox below.

Textbox: Factors affecting metal toxicity

1. properties of DOC,
   - the number of binding sites
   - the heterogeneity of binding sites
   - metal binding to humic and fulvic acids
   - the percentage active fulvic acids
2. binding properties of the biotic ligand
3. intrinsic sensitivity of the species
4. other abiotic factors (excluded in BLMs)
   - mixture toxicity
   - other ligands like nitrates and phosphates
   - other competitors like Fe, Al and Mn
5. other biotic factors (excluded in BLMs)
   - food quality or quantity
   - life history
   - etcetera

During conceptual calibration of field studies, the deviations between predicted and observed toxicity were reduced by adjustment of DOC properties, intrinsic species sensitivity or binding properties of the biotic ligand (Chapter 3). Part of the deviation was caused by other factors, other than BLM-parameters, such as mixture toxicity, and food quantity. In this thesis those uncertainties were assigned to the species sensitivity parameter F_{BL}.

Validation of chemical speciation models

The binding of metals to dissolved organic matter has a significant impact on free ion activities (Tipping et al., 2011). The role of DOC as main descriptor for HC5s of Cu, Ni and Zn was quantified in Chapter 5. The conceptual validation of speciation models required reliable measurements of free ion activities of metals and their complexes. This required advanced equipment and proper calibration. It was beyond the scope of this thesis to review all the methods that are available. However it is useful to be aware that different methods for measuring free ion activities exist, and that their outcomes differ (Sigg et al., 2006).

A difference of a factor 5 in Cu-binding capacities of various sources of DOC was reported (Frimmel et al., 1999). Cu activities in natural water samples were higher than predicted activities with WHAM V (Dwane et al., 1998). An overview of validation studies (Chapter 3) showed that active fulvic acid (AFA) was the most common calibration parameter. Best match between WHAM-calculated and measured Cu-activity was obtained when 40 to 80% of the DOC was considered to be AFA and when the rest was considered to be inert for copper complexation. For pragmatic reasons 50% AFA was used in this thesis, which was supported by several other studies (De Schamphelaere et al., 2002a; De Schamphelaere et al., 2002b; De Schamphelaere et al., 2003).

Validity of competitive binding model

The competitive binding concept was coupled to aquatic toxicity in a gill surface interaction model (GSIM), which was later also incorporated in BLMs (Pagenkopf, 1983). The model was based on previous findings and assumptions: 1) that acute toxicity of metals was caused by alteration of the gill function, leading to dis-functioning of the respiratory system, 2) that some metal species were more toxic than others, 3) that metal species form complexes with the gill surface, 4) that the rate of this process is low compared to the test duration, 5) that gills have a finite binding capacity, and 6) that other cations interact with the metal for binding to the gill (Paquin, 2002a). GSIM was able to explain binding of single metals and additive toxic effects of mixtures of Cu, Cd and Zn.

The validity of the GSIM was determined in a straightforward way by chemical analysis of metals in the gills. The principles of the GSIM were adopted to describe toxicity of metals in gill-less organisms too, for example plants, crustaceans and snails.

Extrapolation of BLMs across species

The primary effect of metals found across a wide range of aquatic organisms was the disturbance of cellular ion levels (Paquin et al., 2002a). This phenomenon and the fact that the affinity of the binding of ions to the gills seemed to be rather constant across species, justified the extrapolation of BLM across species. Because it is impossible to derive BLMs for all species in the ecosystem, some species are used as representatives for a larger group of species, with similar phylogenetic properties. *Daphnia magna* is a commonly used test organism, and daphnia BLMs were shown to be valid for other crustaceans too. The accuracy of the *Daphnia magna* BLMs was verified for rotifers (Cu, Ni and Zn), snails (Ni and Zn) and insects (Ni) (De Schamphelaere et al., 2006; De Schamphelaere et al., 2010; Schlekat et al., 2010). BLMs developed for rainbow trout were applied for all other fish species. Even the acute Cu-BLM of *Daphnia magna* was able to predict chronic toxicity to fish *(ECI, 2008)*. In Chapter 3, Co toxicity for *D. magna* could be described by a BLM for *Enchytraeus albidus,* assuming an equal ligand binding affinity. In Chapter 4, BLMs were used to extrapolate toxicity data of many species to site-specific conditions. The extrapolation of a complete SSD showed that the most sensitive species or even taxonomic groups in ecosystem differed amongst sites (See Figure 4.1), which was a start to understand and manage site-specific effects.

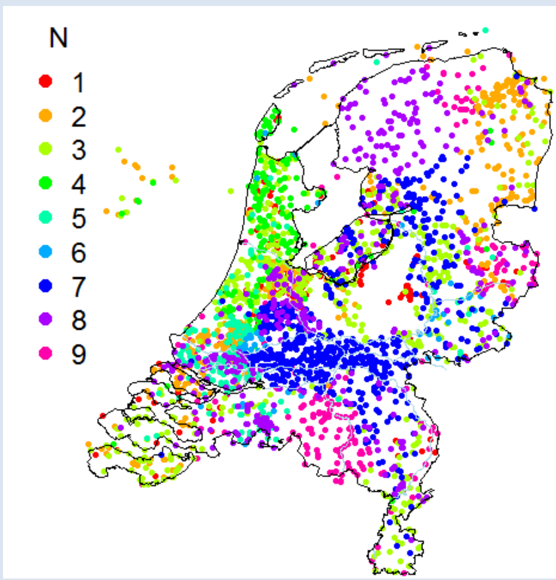
Conclusions and recommendations about BLM validity

Metal concentrations in our study area belonged to the highest found in The Netherlands. Still, effects to *D. magna* were only found on a few sites. BLMs were suitable to rank sites with respect to effects of metals on population growth of *D. magna* (Chapter 3). BLM calibration was needed to predict absolute effect levels, because the presence of other stressors, toxicants and sub-optimal growth conditions affected the species sensitivity. In Chapter 3 we optimized the predictive power of Co, Ni, Zn and Cd BLMs, by calibration of the intrinsic sensitivity or by the introduction of a calibration factor of 20-30. This is justified because experimental, interclonal and interspecies differences affected species sensitivity.

The calibration factors in this study are applicable to surface waters with relatively high metal concentrations, typical for contamination related to zinc industry. More tailored research is required to find out if the calibration factors are relevant for areas predominantly influenced by other sources, land use or geochemical origin. Selection of sites could be based on overlays of maps with total dissolved metal concentrations and for instance land-use maps.

Dealing with data gaps

In most European countries, including the Netherlands, metal concentrations in surface water are regularly monitored by national and regional water authorities.



	n
Total unique sites	4890
pH	3276
Ca	2789
Mg	2123
Na	2135
Cl	4493
SO4	3789
HCO3	2156
DOC	569
DOC+Ca	252
DOC+Ca+pH	228
Complete BLM set	119

Figure 6.3 Coverage of BLM parameters measured on individual sites in the Netherlands. N = the number of BLM parameters measured for particular sites. In the table the number of sites (n) for each BLM parameter is summarized. (source: Limnobase 2007-2010).

Risk assessment of metals, based on total dissolved metals, showed that many sites do not comply with current environmental quality criteria (Chapter 1). On the other hand, many studies showed that the risks were much lower, caused by the mitigating effect of DOC and hardness on metal bioavailability. The risk assessment was refined by BLM calculations which confirmed that risks of Cu and Zn were generally lower, though risks of Ni could be larger than expected (Table 4.2, page 78), and the locations at risk changed.

Because BLM account for the effect of local water chemistry on bioavailability of metals, they are valuable tools for risk managers to identify and prioritize sites that are at risk. Frequently data gaps were present in existing and BLM equations could not be performed. Simplified equations (Chapter 5) optimized the data use and required only a maximum of 3 out of 9 BLM parameters. Moreover, these equations were able to mimic BLM computations with sufficient reliability.

The usefulness of an existing national database over 2007-2009, increased from 2.4% to 12% when a simplified equations with DOC as primary BLM parameter was applied (See Figure 6.3). When aiming for bioavailability corrections with BLMs, future monitoring programs should be optimized to include the required BLM parameters DOC, pH and Ca, and redundant parameters could be removed for this purpose.

Optimal Sampling Frequency

Seasonal variation in risks as a result of variable water composition was observed (Chapter 4). The risk characterization ratio (RCR) is expressed as $PEC/PNEC$. The month of May gave the best reflection of the annual average risk, whereas February exhibited the largest risks. This is a general pattern that could be abstracted when data of 76 sites over a period of 42 months were combined. Monthly risks were approximately 1.19× higher than the annual mean for Cu, for Ni this 1.33× and for Zn 1.85×. This seasonal pattern is quite consistent across sites, though the amplitude of risks may be higher on individual sites (see Figure 6.4).

It is unknown whether the seasonal pattern is generally applicable to other regions. It may be dependent on water temperature (which affects the time that plants start growing, which on its turn changes the water chemistry by uptake of nutrients) and the hydrological situation (is the surface water fed by rain and surface run-off, by melting water or by groundwater). As long as the most vulnerable period is not identified, a general recommendation for the sampling frequency can be derived by a prospective power analysis. A prospective power analysis requires definition of type I (α), type II errors (β), the magnitude of the effect parameter (Δ), and information on the variation from previous investigations.

RCR was used as the effect parameter because it compares the variability in HC5 and metal concentrations. When testing RCR, the effect size was the minimum difference in RCR compared to a reference RCR, that can be proven at the given significance level.

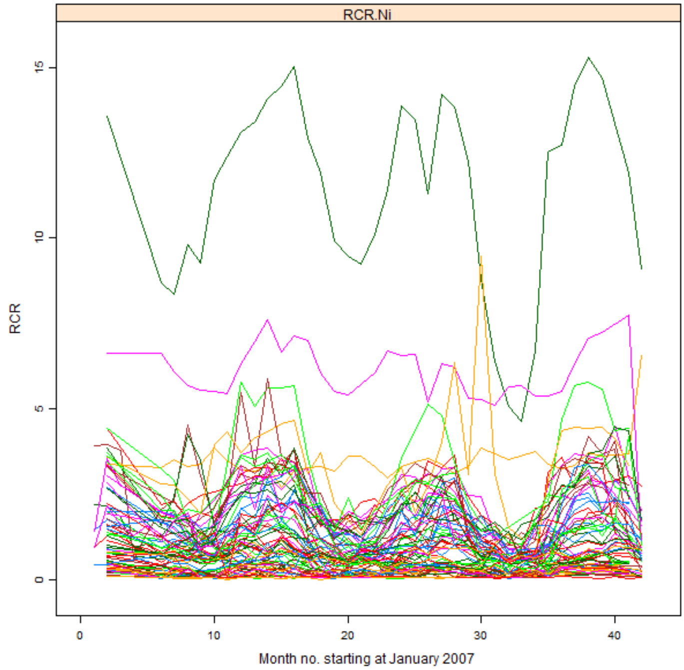


Figure 6.4 Temporal variation of Ni risk in 76 sites in Dommel area

The power analysis was performed, with the aim to detect sites with RCR <0.5 (safe) and sites with RCR>2 (at risk). As a consequence, the risks of sites with 0.5<RCR<2, remains undecided.

Figure 6.5 shows that each sites had its own optimum sampling frequency (*n)* and that differences existed between the metals Cu, Ni and Zn. The sites were ranked, with increasing *n*. A common sampling frequency used by water authorities is 12 samples per year (monthly). Figure 6.5 showed that a frequency of 12 samples per year is sufficient for Cu and Ni at most of the sites (95%) for a reliable RCR estimation within a factor 2. For Zn, a monthly sampling regime will only give reliable results for 30-35% of the sites.

The optimal sampling frequency for Cu and Ni varies between 3 and 21 samples per year. For a reliable assessment of Zn-risks considerable higher sample numbers are required: upto 68 samples per year. High sampling frequencies may not be feasible, from logistic or financial point of view. We therefore recommend that sites with high optimal frequencies (for example >12 times/year) are critically evaluated. The optimal sampling frequency can further be reduced by taking into account the expected concentration range, based on historical data.

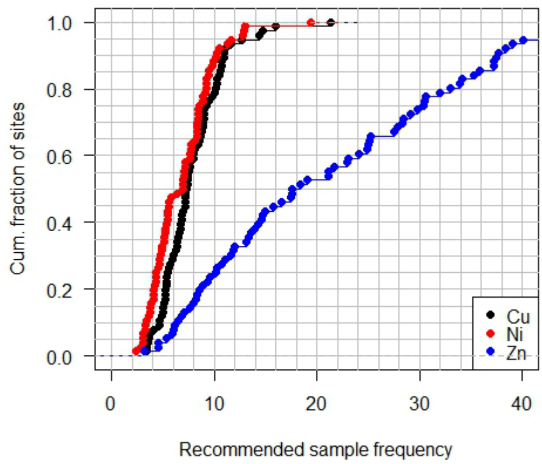


Figure 6.5 Calculated sampling frequency for Cu, Ni and Zn risks, ranked for 76 individual sites. The sampling frequency should be able to detect HC5 exceedances by a factor 2. The sampling frequency is computed, using the observed standard deviation on each site, at α=0.05, β=0.05. In order to meet the assumptions of the power test, RCR values were log-transformed prior to the analysis. In that way, RCR behaves more like a normal distribution, and standard deviation becomes independent of RCR.

Safety factors

It is common practice in current risk assessment methodology, to apply safety factors that account for different types of variations and uncertainties (EC, 2011):

1. overall quality of the database and the endpoints,
2. diversity and representativity of the taxonomic groups,
3. mode of action data,
4. statistical uncertainties around the HC5 estimate,
5. effects data from the field.

In this paragraph we elaborate on “Statistical uncertainties around the HC5 estimate” (EC, 2011) by quantification of the spatial and temporal variations shown in Chapter 4. For the metals Cu, Ni, Zn an SSD-approach is used to derive the EQS, and safety factors of 1-2 for these metals are considered in the EU (Denmark, 2008; EC, 2010; ECI, 2008). The spatial and temporal variation of metal bioavailability determined in Chapter 4 exceeded safety factors that are currently applied for Cu, Ni and Zn. A basic principle of extrapolation in environmental risk assessment is that, where uncertainty is high, larger safety factors are necessary. The magnitude of a safety factor depends primarily on the number and quality of the available toxicological data (See Table 6.1).

Table 6.1 Magnitude of the safety factor, dependent on data availability (EC, 2011)

Available data	**Safety factor**
At least one short-term L(E)C50 from each three trophic levels (fish, invertebrates (preferred *Daphnia*) and algae) (i.e. base set)	1000
One long-term EC10 or NOEC (either fish or *Daphnia*)	100
Two long-term results (e.g. EC10 or NOECs) from species representing two trophic levels (fish and/or *Daphnia* and/or algae)	50
Long-term results (e.g. EC10 or NOECs) from at least three species (normally fish, *Daphnia* and algae) representing three trophic levels	10
Species sensitivity distribution (SSD) method	5-1 (to be fully justified case by case)
Field data or model ecosystems	Reviewed on a case by case basis

For small datasets safety factors vary from 10-1000 and reflect uncertainty due to the low number of species involved and the extrapolation from acute to chronic exposures. When only a few data are available, the lowest EC50 or NOEC is selected and divided by the safety factor to obtain the EQS (EU 2011). For metals also bioavailability or natural background concentrations may be taken into account, to determine the magnitude of the safety factor.

Natural variations in metal-induced risks can be expressed as an uncertainty factor UF (Ragas et al., 1999), which is the ratio between the mean value and a realistic worst-case value (RWC):

$$UF=MEAN/RWC$$

A realistic worst case, is represented by a relatively low HC5, but is not the most extreme values. Therefore 5th percentiles were selected to represent realistic worst case.

In our datasets, the spatial uncertainty was higher than temporal uncertainty (see Table 6.2). Overall, the variability of Cu-HC5 is approximately twice as high as Ni and Zn HC5. The uncertainty factors can be employed on the site-specific HC5, when only limited monitoring data are available, and vulnerable conditions are not defined.

A safety factor for the site-specific HC5, which accounts for spatial and temporal variation, is not necessary as long as monitoring and risk assessment are directed to the most vulnerable conditions. Spatial and temporal variation in HC5 can be accounted for by BLM or by the simplified functions (Chapter 5). When only limited monitoring data are available and not relevant for the most vulnerable conditions, a default safety factor of 5, as recommended by the TGD (EC, 2011), is a reasonable estimate for the combined spatial and temporal uncertainty of Ni and Zn, but for Cu a safety factor of 10 would be more appropriate.

Table 6.2 Uncertainty factors due to spatial and temporal variation in HC5. Spatial variation is based on annual average data per site, temporal variation is based on 42 measurements per site over the period 2007-2010. Overall uncertainty is the product of spatial and temporal uncertainties.

	Spatial	Temporal	Overall
Cu	4.5	2.6	11.7
Ni	5.4	1.2	6.5
Zn	3.9	1.5	5.9

The desired protection level of aquatic ecosystems is ultimately a political choice. The value for the safety factor must be considered in combination with other choices and assumptions in the derivation of HC5, in order to prevent an attenuation of safety factors and other worst-case assumptions.

The power of biotic ligand models lays in their ability to predict water type-specific no-effect levels and metal HC5s. This thesis showed that, after adjustment of the sensitivity factor, BLMs were able to predict effects of multimetal exposure to *Daphnia magna.*

Uncertainties and variations in risk assessment were explicitly shown by application of BLMs to Dutch surface waters. BLMs are able to identify vulnerable time periods and vulnerable species in a particular water type. By knowledge of the seasonal variations of HC5 and metal concentrations, the monitoring frequency could be optimized. Accurate HC5 predictions were obtained with simplified equations, allowing reduction of the number of required monitoring parameters. The consideration of uncertainties and variations offers valuable information on where, when and how to invest in improvement of surface water quality.