National Institute
for Public Health
and the Environment

**Report 607013010/2009**
A.M. Durand et al.

# Toxicity measurements in concentrated water samples

## Evaluation and validation

rivm

RIVM Report 607013010/2009
Centre for Water Management Report 2009.003

# Toxicity measurements in concentrated water samples

Evaluation and validation

A.M. Durand
S. Rotteveel, Centre for Water Management, Lelystad
M.T. Collombon, (present address: Bureau Waardenburg)
E. van der Grinten
J.L. Maas, Centre for Water Management, Lelystad
W. Verweij

Contact:
Anke Durand
Laboratory for Ecological Risk Assessment
anke.durand@rivm.nl

This investigation has been performed by order and for the account of Ministry of Housing, Spatial Planning and the Environment, within the framework of project M/607013 'Bioassays' and of the Directorate-General for Public Works and Water Management, within the framework of project C243.0003 'Bioanalyse'.

rivm

# Abstract

**Toxicity measurements in concentrated water samples**
Evaluation and validation

The National Institute for Public Health and the Environment, together with the Centre of Water Management (formerly called Institute for Inland Water Management and Waste Water Treatment), developed a method to measure the effects of toxic substances in surface water. This method can be used to find out if, when ecological targets following from the European Water Framework Directive are not met, this is caused by toxic substances. The method can also be used to identify *sources* of toxic substances. Traditionally, mainly chemical techniques are applied for these purposes, but the disadvantage is that they do not cover the large amounts of chemicals potentially present in surface waters.

The method uses bioassays. For this purpose, the response of five species in the water under investigation is studied. If a response is found, the substances causing this can be identified (if desired). The method enables to investigate the *combined* effects of substances (synergistic or antagonistic).

Another advantage can be achieved by sample pre-treatment with a resin, allowing to concentrate the toxic substances. This way acute tests can be used instead of chronic test (which are more expensive) even if no acute effects can be found in the original (not concentrated) sample. In addition, natural factors that influence the toxicity of surface water do not disturb the method after this pre-treatment. Unfortunately, the toxicity of metals can not be investigated with this method.

Key words:
bioassays, toxicity, extraction, sample treatment, validation, toxic pressure

rivm

# Rapport in het kort

**Metingen van toxiciteit in geconcentreerde watermonsters**
Evaluatie en validatie

Het RIVM heeft met de Waterdienst van Rijkswaterstaat (voorheen RIZA: Rijksintituut voor Zoetwaterbeheer en Afvalwaterzuivering) een alternatieve methode ontwikkeld om de effecten van giftige stoffen in oppervlaktewater te meten. Met deze methode is eenvoudig te achterhalen of toxische stoffen de oorzaak zijn als ecologische doelen, die voortvloeien uit de Kaderrichtlijn Water, niet worden gehaald. Ook is de methode geschikt om bronnen van toxische stoffen te identificeren. Van oudsher worden hiervoor vooral chemische technieken ingezet. Die hebben als nadeel dat ze slechts een klein deel van het grote aantal chemicaliën in oppervlaktewater kunnen meten.

De methode werkt met zogeheten bioassays. Hiervoor wordt de reactie van vijf soorten organismen gepeild op het te onderzoeken water. Bij een reactie kan desgewenst uitgezocht worden welke stof hiervan de oorzaak is. Met de methode is ook het versterkende effect van meerdere stoffen bij elkaar te achterhalen.

Een ander voordeel wordt behaald met een voorbehandeling met hars, waarmee de verontreiniging wordt geconcentreerd. Hierdoor is een korte test even effectief als een langlopende, en dus duurdere test, ook als het water niet acuut toxisch is. Natuurlijke factoren die de toxiciteit van water beïnvloeden zijn bovendien met deze methode uitgeschakeld. Wel blijft de toxiciteit van metalen en een beperkt deel van de organische stoffen buiten beeld.

Trefwoorden: bioassays, toxiciteit, extractie, opwerking, validatie, toxische druk

rivm

# Contents


| | | |
|---|---|---|
| **Managementsamenvatting** | | **11** |
| **Summary** | | **13** |
| **Samenvatting** | | **15** |
| **1** | **Introduction** | **17** |
| 1.1 | Scope | 17 |
| 1.2 | Principle of the method | 18 |
| 1.2.1 | The concentration procedure | 18 |
| 1.2.2 | Toxicity testing | 19 |
| 1.2.3 | Interpretation of results | 19 |
| 1.3 | Structure of the report | 21 |
| **2** | **Sampling** | **23** |
| 2.1 | Materials and sampling methods | 23 |
| 2.1.1 | Materials | 23 |
| 2.1.2 | Sampling methods | 23 |
| 2.1.3 | Additional sampling | 24 |
| 2.1.4 | Other factors of concern | 24 |
| 2.2 | Transport and sample storage | 24 |
| **3** | **Concentration procedure** | **27** |
| 3.1 | Introduction | 27 |
| 3.2 | Brief description of the procedure | 27 |
| 3.2.1 | Extraction | 27 |
| 3.2.2 | Elution | 28 |
| 3.2.3 | Distillation | 28 |
| 3.2.4 | Replenish with test medium | 29 |
| 3.3 | Effectiveness of the method | 29 |
| 3.3.1 | Extraction from the water sample | 30 |
| 3.3.2 | Elution from XAD | 34 |
| 3.3.3 | Kuderna-Danish distillation | 35 |
| 3.4 | Performance of the concentration procedure | 36 |
| 3.4.1 | Recovery of various chemicals | 37 |
| 3.4.2 | Influence of humic acid | 42 |
| 3.4.3 | Influence of XAD/ water volume and extraction duration | 44 |
| 3.4.4 | Influence of storage time | 46 |
| 3.4.5 | Additional remarks | 46 |
| 3.5 | Recommendations | 48 |
| 3.6 | Conclusions | 50 |

| | | |
|---|---|---|
| **4** | **Toxicity testing** | **51** |
| 4.1 | Choice of test organism | 51 |
| 4.2 | Selected bioassays | 51 |
| 4.2.1 | Bacteria | 53 |
| 4.2.2 | Algae | 53 |
| 4.2.3 | Invertebrates | 54 |
| 4.3 | Performance parameters for bioassays | 58 |
| 4.3.1 | Precision of the bioassays | 58 |
| 4.3.2 | Sensitivity/linearity/working range | 60 |
| 4.3.3 | Additional remarks | 60 |
| **5** | **Interpretation of results** | **61** |
| 5.1 | Introduction | 61 |
| 5.2 | Judgement of single species | 61 |
| 5.3 | Application of statistical extrapolation techniques | 62 |
| 5.3.1 | Introduction | 62 |
| 5.3.2 | Calculating toxic pressure | 64 |
| 5.4 | Evaluation | 66 |
| 5.4.1 | General | 66 |
| 5.4.2 | Acute to chronic ratio | 66 |
| 5.4.3 | Judgement of results | 67 |
| 5.4.4 | Use of statistical techniques | 69 |
| **6** | **Validation of the entire method** | **73** |
| 6.1 | Introduction | 73 |
| 6.2 | Performance parameters | 74 |
| 6.2.1 | Standardization of the method | 74 |
| 6.2.2 | Precision of the entire method | 74 |
| 6.2.3 | Specificity/selectivity | 76 |
| 6.2.4 | Ruggedness | 77 |
| 6.2.5 | Limit of detection (LOD) and limit of quantitation (LOQ) | 78 |
| 6.2.6 | Measurement uncertainty | 78 |
| 6.3 | Interpretation of the results | 79 |
| 6.3.1 | Ruggedness of pT | 79 |
| 6.3.2 | Repeatability of pT | 80 |
| 6.4 | Summary | 81 |
| **7** | **Applications of the method** | **83** |
| 7.1 | Whole Effluent Assessment | 83 |
| 7.2 | Monitoring | 83 |
| 7.3 | Identifying sources | 86 |
| 7.4 | Effect monitoring in water management | 89 |

| | | |
|---|---|---|
| **8** | **Discussion** | **91** |
| 8.1 | Introduction | 91 |
| 8.2 | General | 91 |
| 8.2.1 | Detection additional effects by measuring additional endpoints | 91 |
| 8.2.2 | Ecological relevance | 92 |
| 8.3 | Validation | 93 |
| 8.3.1 | Optimisation concentration procedure | 93 |
| 8.3.2 | Approach further validation | 94 |
| 8.3.3 | Interpretation of the results | 94 |
| 8.4 | Evaluation of applications of the method | 95 |
| 8.4.1 | Toxic pressure versus toxic potency | 95 |
| 8.4.2 | Usefulness for WFD | 96 |
| 8.4.3 | WEA application | 97 |
| 8.5 | Conclusions | 97 |

**List of abbreviations** **99**

**References** **101**

**Appendix I. Protocols** **107**

**Appendix II. Details on chemicals: substance properties and CAS-numbers** **109**

**Appendix III. Calculation of the toxic pressure of metals** **111**

rivm

# Managementsamenvatting


Dit rapport behandelt bioassays, een alternatief voor chemische technieken om de waterkwaliteit te meten. Van oudsher worden vooral *chemische* technieken ingezet voor monitoring van oppervlaktewater. Daaraan is echter een aantal nadelen verbonden. Het aantal chemische stoffen is zeer groot. Alleen al het aantal in Europa geregistreerde stoffen die door de industrie geproduceerd worden is meer dan honderdduizend (http://ecb.jrc.ec.europa.eu/esis/index.php?PGM=ein, website van de Europese Commissie). Dan is er nog geen rekening mee gehouden dat bij de productie van een stof vaak andere stoffen ontstaan, noch dat stoffen die in het milieu komen vaak omgezet worden in andere stoffen. Hoewel niet al deze stoffen in water terecht zullen komen, is het aannemelijk dat het aantal chemicaliën in oppervlaktewater in de tienduizenden loopt.
Het is duidelijk dat het niet doenlijk is al deze stoffen analytisch-chemisch te meten. Maar de onvermijdelijke vraag is dan: welke wel en welke niet? Die vraag is erg lastig te beantwoorden, ook omdat er steeds nieuwe stoffen als 'probleemstof' worden aangeduid. Bovendien levert informatie over een concentratie nog geen inzicht in het *effect* van een stof.
Dit rapport gaat over een alternatief voor chemisch meten, namelijk bioassays. Dat zijn technieken waarbij organismen (bijvoorbeeld algen of watervlooien) worden gebruikt om de *effecten* van toxische stoffen te meten. Bioassays hebben als voordeel dat daarmee in principe de effecten van *alle* stoffen kunnen worden gemeten, inclusief eventuele combinatie-effecten.

Bioassays kunnen op verschillende manieren worden ingezet. De methode die hier wordt beschreven bestaat uit

- een extractiestap waarin de toxische stoffen op een hars worden geconcentreerd;
- een aantal stappen om de stoffen over te brengen in een reeks waterige oplossingen met verschillende concentraties.

Deze manier heeft als voordelen:

- dat resultaten van meerdere bioassays in één getal kunnen worden uitgedrukt;
- dat ook als er geen acute effecten van stoffen meetbaar zijn in het oppervlaktewater, er toch een resultaat kan worden vastgesteld;
- dat zelfs als er geen chronische effecten meetbaar zijn, het mogelijk is om het risico op toxische effecten van verontreinigingen te kwantificeren;
- dat eigenschappen van het water die niet direct door toxische stoffen worden veroorzaakt (bijvoorbeeld pH, zoutgehalte) geen invloed hebben op het eindresultaat.

Nadelen zijn dat niet alle stoffen even goed worden meegenomen bij de gehele procedure. Met name metalen en een (beperkt en bekend) deel van de organische stoffen blijft buiten beeld. Ook is de methode soms lastig te overzien doordat die uit meerdere, soms complexe, stappen bestaat. Verder bestaat altijd het risico dat tijdens de bewerkingen onbedoeld toxische stoffen worden geïntroduceerd.

In de afgelopen jaren zijn allerlei aspecten van de methode onderzocht en gepubliceerd in rapporten en artikelen. De bedoeling van dit rapport is alle kennis bij elkaar te brengen. Tijdens het maken van dit rapport bleek dat een aantal aspecten nog niet goed onderzocht was. Geprobeerd is zoveel mogelijk ontbrekende kennis nog tijdens het schrijven te vergaren om dit rapport zo compleet mogelijk te maken.
In dit rapport wordt onder andere ingegaan op de prestatie-karakteristieken van de techniek, zowel van de verschillende onderdelen als van het geheel. Daarnaast worden enkele toepassingen behandeld. Een van de voorbeelden is hoe voor de Kaderrichtlijn Water nagegaan kan worden of toxische stoffen een rol spelen in het niet halen van de ecologische doelen.

rivm

# Summary


**Introduction**
Chemical methods are usually applied for investigating water quality. However, the amount of substances that occur in surface water is huge. No chemical-analytical methods are available for many substances and if they are, detection limits are often too high. As a result, it is often difficult to use only chemical measurements to assess toxicological risks for an ecosystem. Bioassays may be a good alternative. One advantage is that the effects of *all* relevant substances can be assessed, including combined effects, should they occur.

For years, effects in surface water have been monitored in Dutch large waters. As the water quality has improved in recent decades, the existing toxicity tests showed hardly any acute effects. Monitoring *chronic* effects however, is expensive. For that reason, the National Institute for Public Health and the Environment (RIVM) in cooperation with the National Institute for Inland Water Management and Wastewater Treatment (RIZA; currently named Centre of Water Management) developed a method to determine the toxicity in surface waters in a relatively cheap way. This method consists of a procedure for extracting and isolating the toxic fraction from surface water, enabling to measure the effect of toxic substances on organisms of different trophic levels. The degree of toxicity enables us to estimate the influence of toxic substances on the aquatic ecosystem.

Implementation of the method, e.g. as part of monitoring for the European Water Framework Directive, requires that the method is described satisfactorily. This implies that there is a need for information on technical specifications (Roig et al., 2003; Quevauviller, 2008). This report contains a description of the method and an evaluation of information that was either published before or new. Presented here is information about the technical specifications of the individual steps as well as of the entire method. Applications are described, including a discussion of the practical relevance of the method.

**Evaluation and validation**
For the concentration procedure (sample treatment), much information was evaluated, either gathered experimentally or from literature. The method appears to be suitable for many substances, although there are differences in recovery. More specifically:

- The method is especially suitable for hydrophobic substances with a narcotic mechanism of action. Recovery was typically between 88 and 100%.
- The effects of herbicides and organochloro-pesticides (insecticides) can be well demonstrated. Recovery is usually somewhat lower (60 to 75 %) than for substances mentioned under the first bullet.
- Hydrophilic substances (like medicines) do not adsorb very well onto the applied resins (XAD 4/8). Perhaps other types of resins would lead to better results.
- Volatile substances will disappear largely during the concentration procedure.
- Metals are not extracted from the water phase.
- The extraction procedure can be applied well to surface water. Natural organic matter does not appear to influence extraction efficiency of toxic substances.
- The procedure is well described in protocols. Disturbing factors are well known and included in the protocols.
- Reproducibility of the method is good and within acceptable limits.

The following criteria were used to choose appropriate biological methods:

- acute test with whole organisms;

- endpoints mortality, growth inhibition or photosynthetic activity;
- tests can be performed in small volumes;
- can be performed by laboratory staff without specific expertise.

The set of organisms was tuned to obtain optimal ecological relevance, by including primary producers (algae), primary consumers (crustaceans) and decomposing organisms (bacteria). The tests are well described in protocols and mostly derived from international ISO-standard tests. Reproducibility of the test is good and within acceptable limits, both in intra- and in inter-laboratory experiments.

Data interpretation was done in two ways. The first method can be applied to individual tests and uses a criterion for toxicity that is likely to indicate ecosystem effects. The second method applies to the *set* of test and is based on a risk analysis, using a species sensitivity distribution. The methods have in common that they indicate the potential negative influence of toxic substances on ecological status.

Validating the entire method in the same way as is done for chemical methods is complex. The reproducibility of the entire method is good and comparable to the reproducibility of the individual tests. Other technical specifications, such as LOD, LOQ, linearity and precision are difficult to establish because the chemical composition of surface water is not known (can only be established for a single substance or a mixture of known substances). For that reason, no detection limit, linearity and 'bias' are specified. There was, however, a pragmatic solution chosen for LOQ-determination.

**Applicability**
Experiences show that the method is well applicable. Within the European Water Framework Directive it has its added value by pointing to locations where toxic substances may negatively impact ecological quality (diagnosis); for that reason it is proposed as an additional method within WFD-monitoring (Maas, 2005). The method can also be applied for effluent assessment and for determining the effects of additional waste water treatment. Biological effect measurements also prevent unlimited growth of lists of substances to be monitored, which may lead to lower monitoring costs.

**Recommendations**
The extraction of hydrophilic substances should be improved in order to assess new types of relevant substances, including medicines. In order to improve ecological relevance, other representative and more sensitive organisms may be included.

rivm

# Samenvatting


**Introductie**
Waterkwaliteit wordt in veel gevallen met chemische methoden onderzocht. Het aantal stoffen dat in oppervlaktewater voorkomt is echter zeer groot. Voor lang niet alle stoffen zijn meetmethoden beschikbaar, en voor zover ze beschikbaar zijn, zijn de detectiegrenzen vaak te hoog. Daardoor is het soms lastig op basis van alleen chemische metingen vast te stellen of er sprake is van toxicologische risico's in een ecosysteem. Biologische meetmethoden (zogeheten bioassays) kunnen een goed alternatief vormen. Voordeel is onder andere dat de effecten van (in principe) *alle* relevante stoffen worden bepaald, inclusief eventuele combinatie-effecten.

Effecten in oppervlaktewater worden al gedurende langere tijd in de Nederlandse rijkswateren gemonitord. Aangezien de waterkwaliteit de afgelopen decennia verbeterd is werden met de bestaande toxiciteitstesten nauwelijks nog acute (korte termijn) effecten aangetoond. Het monitoren van langetermijneffecten is echter kostbaar. Daarom is door RIVM en RIZA (thans Waterdienst) een methode ontwikkeld, die de toxiciteit van het oppervlaktewater kan bepalen op een kosteneffectieve manier. De methode die daarbij gebruikt wordt berust op het extraheren en concentreren van de toxische fractie uit oppervlaktewater, waarna het effect van de toxische stoffen op organismen van meerdere trofische niveaus gemeten wordt. Aan de hand van de hoogte van de toxiciteit kan een uitspraak gedaan worden over de mate van invloed van toxische stoffen op het aquatisch ecosysteem.

Implementatie van de methode, bijvoorbeeld als een onderdeel voor de KRW monitoring, vereist een goede beschrijving van de methode. Hiervoor is inzicht nodig in diverse prestatiekenmerken (Roig et al., 2003; Quevauviller, 2008) Dit rapport omvat een beschrijving van de methode (met bijbehorende protocollen), waarbij eerder gepubliceerde en nieuwe informatie is geëvalueerd. Van zowel de aparte onderdelen van de methode als van de methode als geheel zijn de prestatiekenmerken (zoals betrouwbaarheid en reproduceerbaarheid) gepresenteerd. Naar aanleiding van diverse toepassingen is ingegaan op de relevantie van de methode.

**Evaluatie en validatie**
Voor de concentratieprocedure is veel uitgezocht, zowel experimenteel als vanuit de literatuur. De methode blijkt geschikt voor een groot aantal stoffen, hoewel de recovery niet voor alle stoffen even goed is. Meer specifiek:

- De methode is met name geschikt voor hydrofobe stoffen met een narcotische werking. Recovery percentages van 88 – 100 % worden behaald.
- De effecten van herbiciden en organochloorbestrijdingsmiddelen (insecticiden) zijn goed aantoonbaar. Recovery percentages liggen over het algemeen wat lager (60 – 75 %).
- Hydrofiele stoffen (zoals diverse medicijnen) adsorberen minder goed aan de combinatie van kunstharsen (XAD4/8) die voor de extractie wordt gebruikt. Wellicht dat anderen typen hiervoor een verbetering opleveren.
- Vluchtige stoffen zullen in de procedurestap grotendeels verdwijnen.
- Metalen worden eveneens niet geëxtraheerd vanuit de waterfase.
- De extractieprocedure is goed toepasbaar op oppervlaktewater. Humuszuren in oppervlaktewater lijken geen invloed te hebben op de extractie efficiency.
- De procedure is goed geprotocoliseerd. Factoren die van invloed kunnen zijn op de procedure zijn goed omschreven en in het protocol vastgelegd.
- De reproduceerbaarheid van de methode is goed en ligt binnen acceptabele waarden.

De biologische methoden zijn geselecteerd op:

- acute testen met gehele organismen;
- eindparameters sterfte, groeiremming of fotosynthese-activiteit;
- testen in minimale volumes;
- goed uitvoerbaar zonder specifieke training.

Bij de keuze van de testbatterij is zoveel mogelijk rekening gehouden met de ecologische relevantie door gebruik te maken van primaire producenten (algen), primaire consumenten (kreeftachtigen) en reducenten (bacteriën). De testen zijn goed geprotocoliseerd, en zijn veelal herleid vanuit internationale ISO-standaard testen. De reproduceerbaarheid van de testen is zowel in intra- als interlaboratorium-experimenten goed en binnen acceptabele grenzen.

De interpretatie van de gegevens is uitgewerkt voor zowel de testen afzonderlijk als voor de batterij aan toxiciteitstesten. In de eerste methode wordt een criterium voor de toxiciteit aangegeven, waarbij verwacht wordt dat in het ecosysteem effecten op kunnen gaan treden. De laatste methode is gebaseerd op een risico analyse, waarbij gebruik gemaakt wordt van een gevoeligheidsverdeling. Beide methoden geven aan wanneer stoffen een relevante rol spelen voor het wel of niet behalen van een goede ecologische kwaliteit.

Het valideren van de gehele methode blijkt lastig te zijn ten opzichte van chemische methoden. De reproduceerbaarheid van de gehele methode is goed en blijkt niet veel af te wijken van de reproduceerbaarheid van de toxiciteitstesten alleen. Overige prestatiekenmerken, zoals de Limit of Detection, Limit of Quantification (LOQ), lineariteit, afwijking ten opzichte van de werkelijk verwachte waarde zijn moeilijk vast te stellen, omdat voor een onbekend mengsel niet duidelijk is welke waarde te verwachten is. Deze waarden zijn af te leiden voor een enkele stof of een mengsel van bekende stoffen. Voor het meten van toxiciteit in een mengsel, zoals oppervlaktewater, is niet bekend om welke stoffen het gaat. In dit rapport is daarom geen detectielimiet, lineariteit en ‘bias’ gegeven. Er is wel een pragmatische invulling gegeven aan het bepalen van de LOQ.

**Toepasbaarheid**

In de praktijk is gebleken dat de methode al goed toepasbaar is. Binnen de KRW heeft het zijn meerwaarde door aan te tonen waar stoffen mogelijk nog een rol spelen voor het bereiken van een goede ecologische kwaliteit (diagnose), en is daarom gepresenteerd als een aanvullende methode binnen de KRW-monitoring (Maas, 2005). Ook is de methode goed toepasbaar bij de beoordeling van emissies en bij het vaststellen of maatregelen bij een zuiveringsstap effect hebben gehad. Biologische effectmetingen omzeilen bovendien het meten van een steeds uitgebreidere lijst van chemische stoffen, wat kostenbesparend kan werken.

**Aanbevelingen**

De extractie van hydrofiele stoffen moet verbeterd worden om aan te kunnen sluiten op nieuwe relevante stoffen, zoals medicijnen. Voor de verbetering van de ecologische relevantie is het wenselijk om de methode aan te vullen met andere representatieve en gevoeliger testorganismen.

rivm

# 1 Introduction

## 1.1 Scope

The chemical quality of Dutch surface waters has improved substantially in recent decades, particularly in the major rivers, though concentrations of several contaminants have not yet been reduced to target values. The question is to what extent the aim of protecting the ecosystem from exposure to contaminants has been achieved. Since there are over 100,000 substances (an amount that is still increasing), the monitored chemicals represent only a small part of the entire range of chemical substances. There is some information on single toxicity of monitored substances, but combined effects from a mixture of substances can also occur. Diagnosis of toxicological stress on ecosystems can help identify areas of toxicological concern or evaluate regulatory actions (Posthuma et al., 2002). Toxicological stress can be measured with bioassays: experiments in which living organisms or parts of organisms (tissues or cells) are exposed to environmental samples to show the biological effects of contaminants.

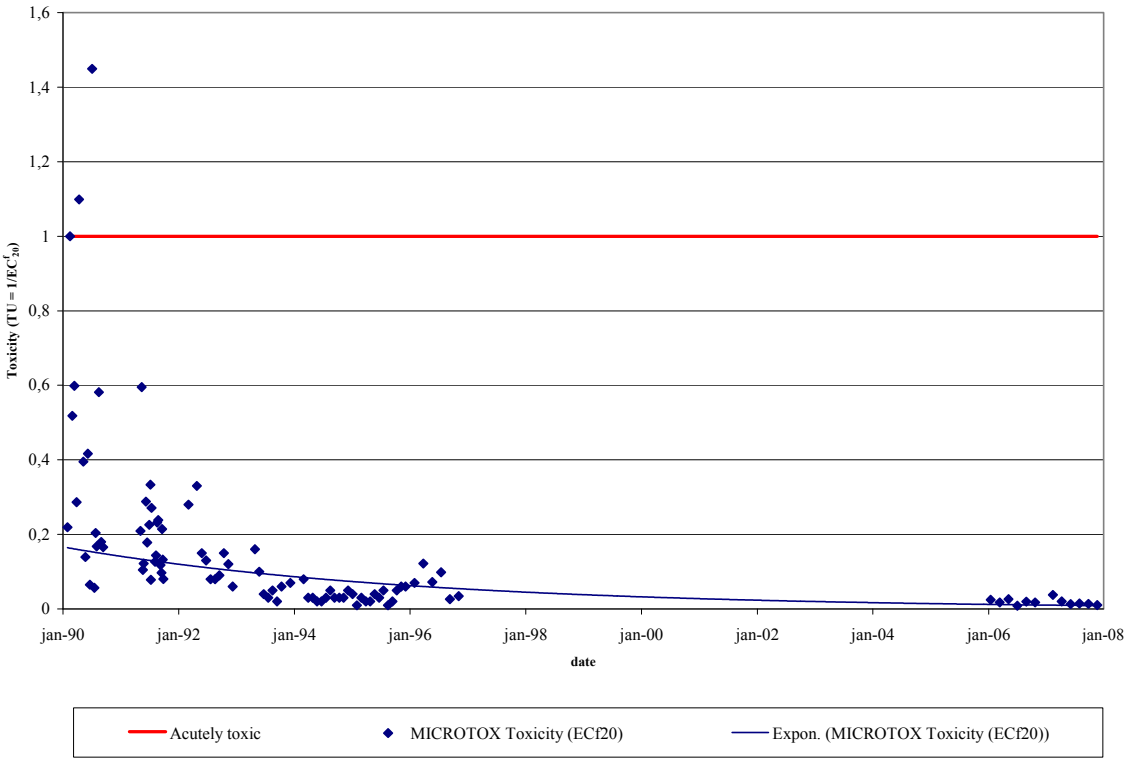


**Figure 1-1. Available tests, like the Microtox® assay (a test with bioluminescent bacteria) in this example, barely show any effects at the present level of contamination as a consequence of the improvement in water quality (De Zwart and Sterkenburg (2002), completed with recent monitoring data).**

Since the early 1990s, water quality in the Netherlands has been regularly assessed by chemical and biological monitoring (RWS, 1999; Maas and Van den Heuvel-Greve, 2005). Bioassays on

environmental samples respond to the combined effect of all substances present in the environment. As a consequence of the improvement in water quality, toxicity tests like the Microtox® assay barely show any acute effects at the present level of contamination (Figure 1-1). Monitoring of environmental toxicity in the range of sub acute or chronic effects is very laborious, and therefore expensive, because of the long time it can take to detect any effect. Furthermore, any effect found is difficult to assess because of the problems of separating the effects of toxic substances and other natural stress factors. With these considerations in mind, the National Institute for Public Health and the Environment (RIVM, Bilthoven) and the former Institute for Inland Water Management and Wastewater Treatment (RIZA Lelystad, currently Centre of Water Management) have developed and optimized a method for measuring toxic pressure in the aquatic system in a cost-effective way.

The method consists of a concentration procedure, toxicity testing and (statistical) interpretation of results. First, the organic toxicants in the water sample are adsorbed to a synthetic resin, eluted with an organic solvent and concentrated into a water extract. Then, the toxicity of the extract is measured using a set of standardized acute tests. The results can be interpreted in a single species approach or in a risk-based manner.

Toxicological stress has been monitored in several projects. Trends in toxic stress have been determined as part of the national rivers monitoring programme (De Groot et al., 2003). The method gives a good indication of the variability in toxic stress in different seasons and a good comparison of toxic stress on locations in a river basin. The possibility of using this method on Whole Effluent Assessment (WEA) is being investigated (Roex, 2005). The method can also be used for monitoring the results of a particular treatment stage at sewage treatment plants (Roex and Rotteveel, in prep.). Internationally, the method has been presented as an alternative or additional tool for monitoring under the Water Framework Directive (WFD; Maas, 2005).

Effect measurements have not been implemented in the WFD monitoring. Bioassays on water extracts could nevertheless provide useful information on the question if the ecological status is at risk due to toxic substances. Implementation of the method, perhaps as an ‘emerging tool’ for monitoring under the WFD, requires a well-defined description of the sampling method, evaluation of the accuracy of the method (precision, limit of detection), calculation of and correction for method bias and an assessment of any uncertainty in the measurements (trueness) (Roig et al., 2003; Quevauviller, 2008) This report presents the method and addresses questions about reliability and reproducibility raised by its application. It therefore combines both new knowledge and background information.

## 1.2 Principle of the method

Besides sampling, the method for estimating the toxicological pressure on a water sample consists of:

a. the concentration procedure;
b. toxicity testing;
c. interpretation of results.

### 1.2.1 The concentration procedure

Physical-chemical techniques are used to isolate and concentrate the organic toxicants from the water sample. This concentration step is necessary to increase the concentration of toxicants, so that acute toxicity tests can be used. At the same time the concentration step has the advantage of separating toxicological stress from other stress factors like minerals and nutrients. The method aims at

comprehensive adsorption to adsorbents with a high affinity for the biologically available fraction of toxic substances. A combination of synthetic resins (called XAD-4 and XAD-8) is used as the adsorbent in the method described in this report (Figure 1-2). Only the organic fraction of the contaminants is involved, since metals do not adsorb to the XAD.

After adsorption, the substances are removed from the XAD using acetone. This solvent appeared to be the most suitable in comparison to other organic solvents. Before using the eluates for toxicity tests the acetone is removed and the substances are transferred to a water extract. This recovery of substances has been investigated for several substances and can be about 80 % for narcotic substances like 3-chloro-nitrobenzene. On the contrary, organotin substances show very low elution efficiencies. A detailed description of the procedure is given in chapter 3, where results from recovery experiments are shown as well.

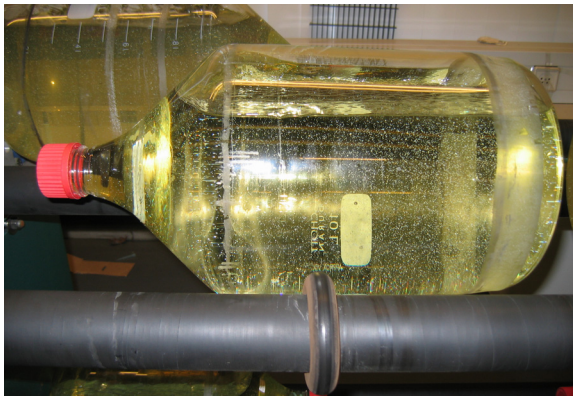

**Figure 1-2. Extraction from substances with XAD. Water and XAD are tumbled for 48 hours on a roller bench.**

### 1.2.2 Toxicity testing

The toxicity of the eluate is measured in toxicity tests (Figure 1-3) using organisms representing different functional groups. Because of the concentration step, the amount of material available is small in relation to the original sample. One factor in the choice of test organisms is that it must be possible to measure the toxicity in very small volumes.
To date, toxicity tests have mostly been performed on whole organisms (in vivo tests). The acute toxicity tests vary in duration from a few hours to several days. A detailed description of the biological tests is given in chapter 4.

### 1.2.3 Interpretation of results

Depending on the aim of the measurements, the results of the bioassays can be interpreted in a single species approach or in a risk-based manner. The results of bioassays can be compared to criteria for effect assessment. These criteria are chosen at the level at which no chronic toxicity is expected and the level at which toxicity is considered negligible.
Two methods for effect assessment are presented:

a. effect indication based on judgement of single species;
b. effect indication using risk analysis based on a species sensitivity distribution.

For less than 4 test species, the single species approach can be used. This method is also used for the assessment of whole effluents. The criteria indicate effects based on exceeding toxicity levels.

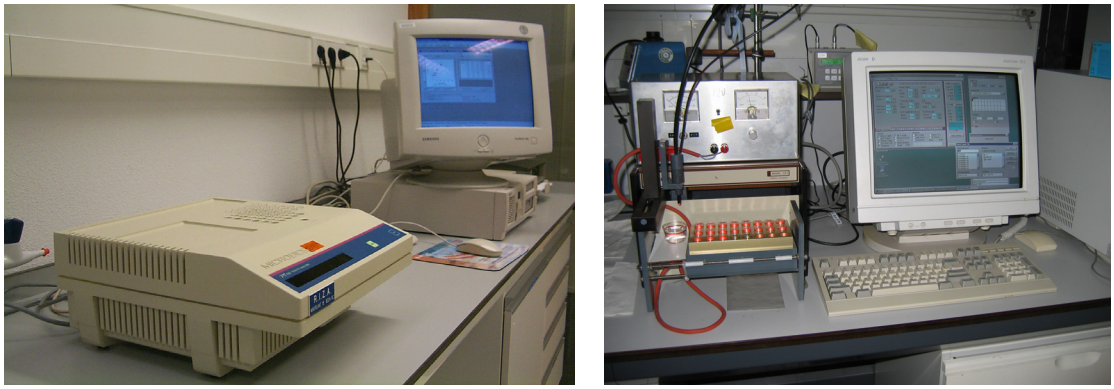

**Figure 1-3. Microtox® test (left) and algae test (PAM, right) which can be applied to very small volumes.**

If more toxicity data are available (four or more test species), a risk analysis can be used based on a species sensitivity distribution (SSD) (Posthuma et al., 2002). The potential fraction affected is derived from the sensitivity distribution at the original sample concentration (Figure 1-4).

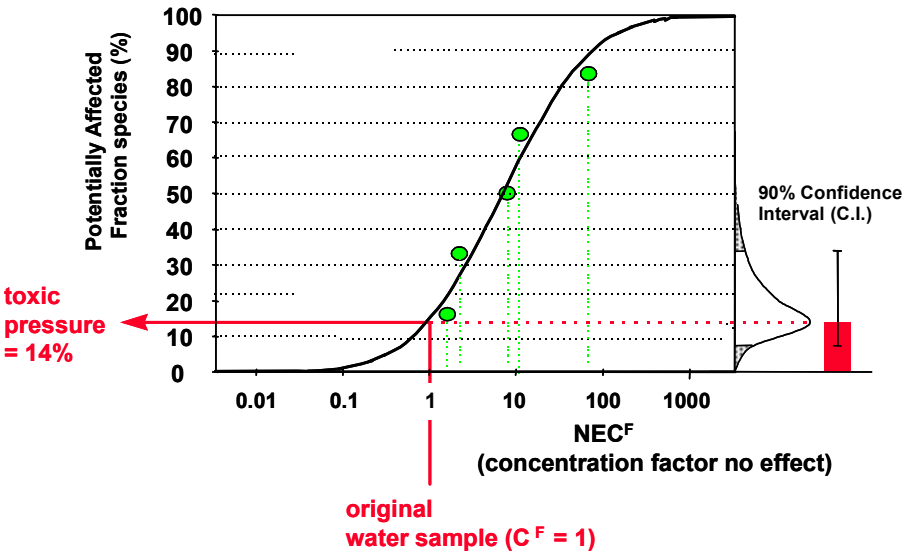


**Figure 1-4. Example of a cumulative species sensitivity distribution curve for the end-points of five bioassay tests, and its extrapolation for the ecological risk in the not concentrated sample (Struijs and De Zwart, 2003).**

Metals are not extracted from the water sample with XAD. However, for most relevant metals, large toxicity datasets are available. These data can be used to calculate a potentially affected fraction (PAF) for measured metals using SSD. Options for calculating toxic pressure for metals are presented in Appendix III.

A detailed description of the interpretation methods is given in chapter 5.

## 1.3 Structure of the report

This report examines the different stages in the method. The sampling procedure and points for attention are discussed in chapter 2. Chapter 3 describes the concentration procedure, discussing the different stages: adsorption, elution and concentration. The choices made and recent developments in the method are also examined. Chapter 4 describes the different bioassays that have been applied in or might be chosen for various applications. In chapter 5 the interpretation of results is examined in a single species approach and a risk-based approach. In chapter 6, the entire method is validated and evaluated. Examples of applications of the method are presented in chapter 7. The relevance of the method is discussed in chapter 8, which is closed off with the conclusion. The Appendices contain protocols for the entire method (from sampling to interpretation), details on chemicals for which concentration procedure is evaluated and the calculation of the toxic pressure of metals. The schematic structure of the report is presented in a diagram in Figure 1-5.

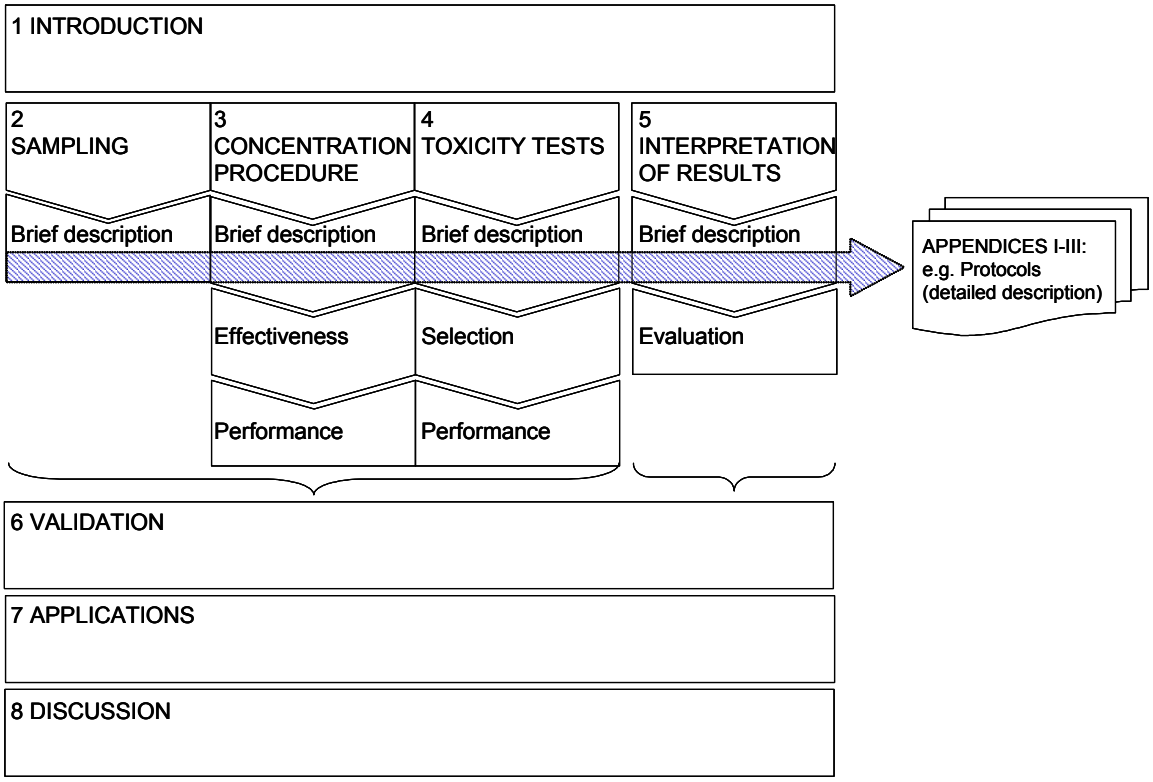


**Figure 1-5. Structure of the report.**

# 2 Sampling

Many factors can influence the chemical composition and toxicity of a sample, including homogeneity, stratification, contamination, adsorption, evaporation, photo-degradation, biodegradation, physical or chemical reactions and hydrolysis. To prevent such negative influences, correct sampling, transport and storage procedures must be followed.

Only the organic substances are isolated using XAD. Additional sampling is needed to be able to measure metal concentrations (see section 2.1.3).

## 2.1 Materials and sampling methods

A large volume of sampling water is needed to retrieve enough concentrate to perform a series of toxicity tests. For example, at least 80 litres of surface water would need to be sampled to assess toxicity in five tests with a thousand-fold concentrated sample (see Appendix I).

### 2.1.1 Materials

Glass containers or stainless steel vessels should be used for organic constituents and polyethylene containers for sampling for metal impurities (ISO 5667- 3 and 6; 2003, 2005). Stainless steel vessels (20-30 litres) are generally used for sampling surface or wastewater. These vessels are most suitable for the transportation of the large sampling volume required.
The use of synthetic materials for sampling and storage should be avoided. Phthalates, used as softeners in synthetic materials, can leach into the sample and influence the toxicity.

### 2.1.2 Sampling methods

Sampling can be performed manually or automatically. These methods, and the materials needed, are discussed separately. For both methods extra caution is needed to prevent contamination due to the sampling, because these contaminations can be concentrated a thousand fold within the concentration procedure. The number of materials used should be kept as low as possible to prevent contamination.

#### Manual sampling

For manual sampling, an open-mouthed vessel (bucket or can) can be used to collect the sample. To prevent contamination, it is recommended that a stainless steel bucket with a rope be used. All materials used for sampling should be rinsed twice with sampling water before use.

The bucket should be filled by immersion (gently pushing) below the water surface. Take care not to scrape the bucket against the boat, bottom or quay when lifting it out of the water. A funnel or pouring cup can be used to pour the water sample into the vessels (20-30 l). While pouring, stir the sample in the bucket regularly to guarantee the homogeneity of the sample. The vessels should be sealed directly after filling. A filter with a wide mesh (100 μm) should be used to avoid intake of coarse particles.

#### Automatic sampling

Automatic sampling can be performed with pumping devices or automatic sampling machines. Several types of pumping devices can be used for automatic sampling, as long as the minimum pumping speed

is sufficient. The system used should be flushed with sampling water for about 5-10 minutes before sampling.

Other automatic sampling devices may be operated on a time-proportional basis. This method is particularly useful in preparing composite samples to study the average substance load in a river or wastewater stream. Contamination of materials (tubes, collecting vessels etc.) should be prevented and to ensure that sample instability does not lead to errors as a result of the longer storage time.

The sample should be taken about 0.5 m below the water surface, with the inlet facing the direction of flow. The distance to the bottom should be at least 0.5 m (if possible). The inlet of the sampling device should be protected by surrounding it with both a coarse and a fine mesh, and should be frequently inspected to ensure the free flow of water is not hindered.

The sampler inlet should be held 1.5 m from the hull if samples are being collected by boat. If a sampling pump is used, the sample can be poured directly into the 20-30 l vessels.

### 2.1.3 Additional sampling

Samples of surface or wastewater taken to estimate the concentration of metals should be collected in polyethylene or borosilicate glass. The samples must be preserved immediately after sampling. For total analysis of metals, the samples must be acidified to pH < 2. For the measurement of dissolved concentrations of metals the samples must first be filtered and acidified to pH < 2.

To calculate the speciation of the metals DOC concentrations are needed. For this parameter 100 ml of surface water is sampled in a polyethylene or glass bottle. The sample could be preserved with acid to a pH of 1-2 and stored at 4 °C or frozen at –20 °C. Samples should be analysed within 7 days or 1 month respectively (ISO 5667-3, 2003).

### 2.1.4 Other factors of concern

Other factors that are important in sampling are:

- the choice of sampling site;
- the importance of mixing (especially in effluent sampling);
- the frequency and timing of sampling.

These factors are referred to in ISO 5667-6 (2005) and apply generally to sampling in rivers and wastewater streams.

## 2.2 Transport and sample storage

After sampling, there are several factors that can influence the toxicity of the sample, so it is important to get the samples to the test laboratory as soon as possible. Low temperature and darkness keep any impact on the samples during transportation to a minimum.

The samples should be transported to the test laboratory immediately after sampling (and at any rate within 24 hours). During transportation and eventual storage, the samples should be kept cold (preferably 4 ºC) and in the dark. Water samples must be treated immediately with XAD after delivery to prevent loss of toxicity.

Storage time should be kept to a minimum to prevent loss of substances. After sampling the water samples should be treated with XAD within 24 hours. In general, organic micropollutants should be

extracted and analysed within 24 hours of sampling (ISO 5667-3, 2003). Metal analyses should be performed within one month if samples are acidified.

# 3 Concentration procedure

## 3.1 Introduction

Concentration of micropollutants is a way of obtaining an indication of toxicological stress in aquatic ecosystems. It allows instantaneous measurement of biological effects; when toxic substances are extracted from water and brought together in a small volume of water, the combined concentrations of toxic substances may be high enough to cause acute toxic effects.

A suitable concentration technique should effectively extract as many micropollutants possible from water, without unintentionally adding toxicity to the concentrate and with a minimum of chemical alterations. In addition, it should be a simple technique that is easy to carry out, so that it can be implemented in routine monitoring programmes if desired.

Therefore, a simple concentration has been developed for concentrating dissolved organic micropollutants from surface water samples into a smaller volume of water. This technique has the advantage of specifically concentrating a wide range of organic micropollutants, while other substances like inorganic salts and humic substances are not concentrated. The concentrated water sample is suitable for toxicity testing. The method can be used for all types of treated or untreated surface water and effluents.

The aim of this chapter is to inform users of the benefits and limitations of the presented concentration method. Developments did not aim at a perfect concentration method, but at a cost-effective method of obtaining information on the toxic pressure in a water system.

This chapter explains the concentration method. First, a brief general description of the principle is given (section 3.2), and then more detailed background information on each part of the method is presented (section 3.3 and 3.4). A detailed protocol is presented in Appendix I-2. In section 3.3 a distinction is made between the factors that principally influence the effectiveness of the method due to choice of materials and methods. The performance of the concentration procedure is discussed in section 3.4.

## 3.2 Brief description of the procedure

In order to be concentrated, substances must be extracted from a large volume of water, and subsequently transferred to a smaller volume of water. The method comprises three steps:

- isolation of organic micropollutants from water by extraction using synthetic resins (XAD);
- removal of the substances from the XAD resins by elution of the resins with acetone;
- replacement of the acetone by water by means of distillation.

### 3.2.1 Extraction

Solid phase extraction (SPE) using a combination of two synthetic resins, XAD-4 and XAD-8, is performed to extract organic micropollutants from water. The purified and conditioned resins are added to an untreated water sample and the water samples are tumbled in order to optimize contact of the

water sample with the XAD (Figure 3-1). After an extraction period of 48 hours, the XAD is collected on a sieve and transferred to a Petri dish, to allow any excess water to evaporate overnight.

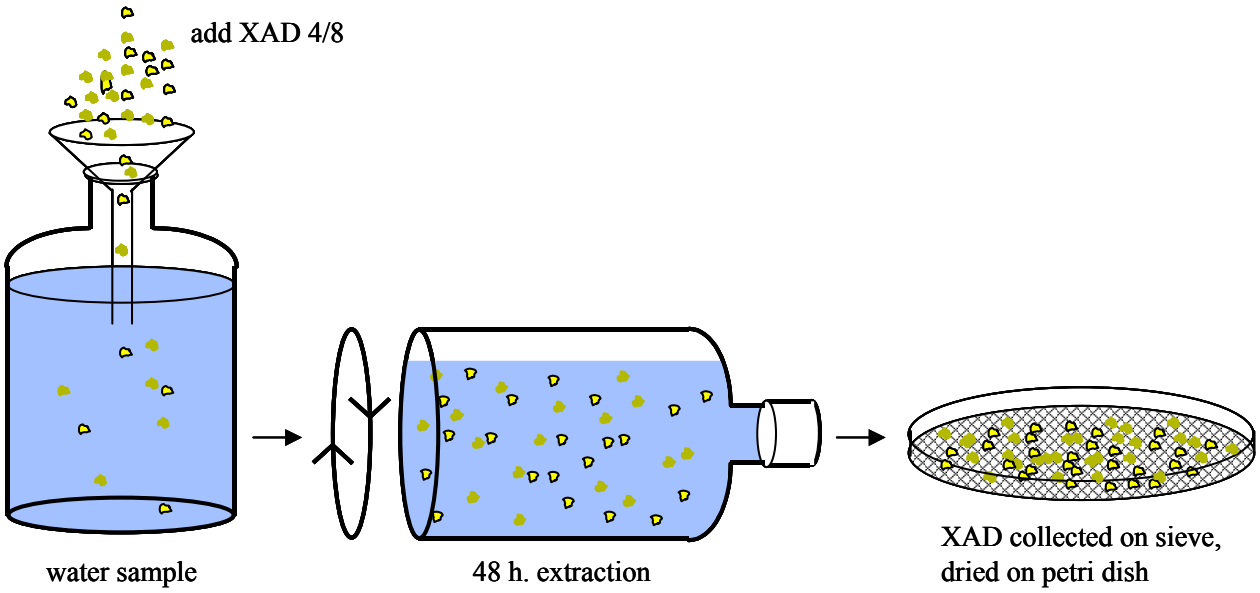


**Figure 3-1. Extraction of substances from water with XAD 4 and 8. After the extraction period the XAD was dried on a Petri dish.**

### 3.2.2 Elution

The sorbed micropollutants are removed from XAD using acetone, resulting in an acetone eluate that contains extracted organic micropollutants and may contain a small amount of water. (Figure 3-2)

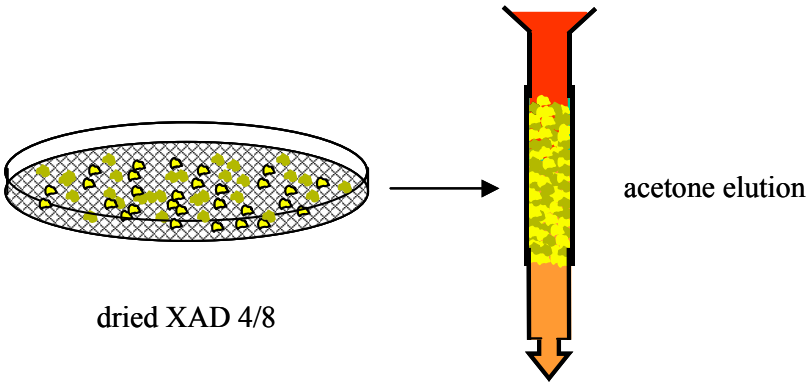


**Figure 3-2. Elution of the substances from XAD with acetone.**

### 3.2.3 Distillation

The acetone is largely removed from the eluate by means of Kuderna-Dänish distillation (Figure 3-3), which takes place at approximately 65 °C.

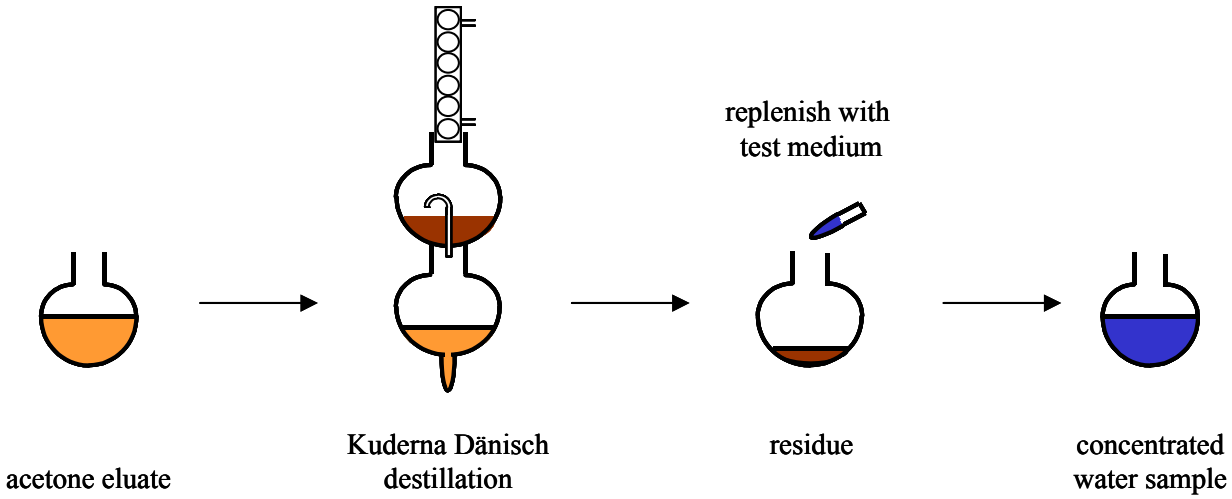


**Figure 3-3. The concentrate was prepared by dilution with medium after removing the acetone.**

### 3.2.4 Replenish with test medium

The residue is replenished with water (EPA medium: Freeman, 1953 and US EPA, 1985) until the desired volume is achieved, resulting in a concentrated water sample that is suitable for ecotoxicity testing (Figure 3-3). The water sample is usually concentrated a thousand-fold for monitoring purposes: 60 litres of water is concentrated to 60 ml. For other purposes it can be chosen to use a lower concentration factor by replenishing the residue to a larger volume.

## 3.3 Effectiveness of the method

When considering application of the XAD concentration method on environmental samples, it may be helpful to have insight into the effectiveness of the method for different kinds of chemicals. In order to be effectively concentrated, the substances must be:

- extracted from water onto XAD;
- eluted from XAD into acetone ;
- non-volatile enough not to evaporate during distillation;
- stable enough not to break down during distillation;
- soluble enough not to exceed maximum water solubility in the concentrated sample.

Concentrating environmental water samples is almost by definition a matter of working with unknown mixtures of micropollutants. Many factors influence the recovery of chemicals at the end of the concentration procedure. It is therefore important to realize that it is impossible to know the exact effect of the different operations in the concentration procedure. Variation in handling of a sample may result in increased recovery variation. For this reason, it is important to standardize the procedure as much as possible.

This section specifies the principles on which the method is based, which gives insight in factors influencing the effectiveness of the method.

### 3.3.1 Extraction from the water sample

The process of extracting micropollutants from water is based on competition between the adsorbing agent and the surrounding water. This competition is affected by:

- the choice and amount of adsorbing agent;
- the composition of the water sample (pH, ionic strength, humic acids and particulate matter);
- substance properties (organic or inorganic micropollutants, hydrophobicity or hydrophilicity, polar substances and steric exclusion) and
- environmental parameters (temperature, light, et cetera).

**XAD properties**

The non-polar polystyrene divinylbenzene copolymer XAD-4 (Figure 3-4) has a large surface area and a pore diameter of 50 Å. XAD-8 is a polymethyl methacrylate resin (Figure 3-4) and is somewhat more polar than XAD-4. The pore diameter is larger and the surface area smaller than that of XAD-4 (Table 3-1). XAD-4 and -8 are resistant to acetone and are stable in pH range 0-14 (Struijs and Van Buren, 1995). Since the applied XAD resins consist of uncharged molecules, they are not able to isolate ionogenic substances.

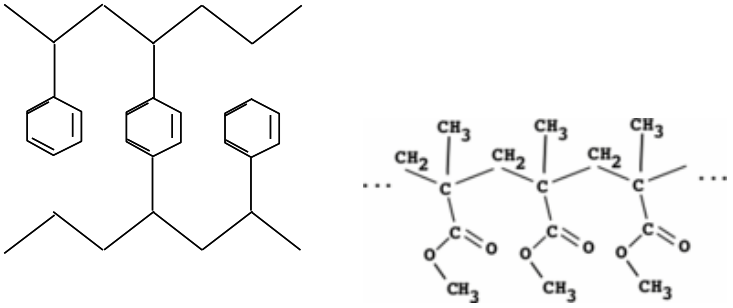

**Figure 3-4. Structures of (left)XAD-4; polystyrene divinylbenzene copolymer resin and (right) XAD8; polymethyl methacrylate resin.**

**Table 3-1. Properties of XAD-4 and XAD-8.**

| | XAD-4 | XAD-8 |
|---|---|---|
| Resin type | Polystyrene divinyl benzene | Polymethyl methacrylate |
| Surface area | >750 $m^2$/g | 160 $m^2$/g |
| Porosity | >0.5 ml/ml | 0.79 ml/g |
| Pore size | 50 Å | 225 Å |
| Bead size | 0.49-0.69 mm | 0.40-0.60 mm |

**Choice for adsorbing agent**

The synthetic resins that are used for extraction of pollutants from water, are organic polymers known as XAD. XAD synthetic resins are hard, insoluble beads of high-surface porous polymer, known as macro reticular resins (see textbox on XAD properties). The combination of the two resins XAD-4 and XAD-8 was chosen based on their known ability to extract a wide range of organic components (Struijs

and Van Buren, 1995). Both resins are stable over a wide range of pH values (AWWA-KIWA, 1988) and are easy to handle.

The amount of XAD relative to the water volume and the extraction duration may influence the recovery of the substances (Struijs and Van de Kamp, 2001). The chosen XAD:water volume ratio will be discussed in section 3.4.3.

**Composition of the water sample**

Several water parameters influence the effectiveness of substance extraction from water onto XAD. The pH value, ionic strength, the presence of humic acids and particulate matter may affect adsorption efficiency.

*pH*

Influence of pH is changing equilibria between dissociated and undissociated forms of molecules in one direction or the other. At pH=7, organic acids and bases that are completely dissociated will not be (readily) adsorbed. At lower pH, organic acids will be extracted to a greater extent, while a higher pH favours organic bases.

*Ionic strength*

The concentration of ions in the water sample can influence adsorption to XAD. Ions with the right charge can neutralize ionized organic components, enabling the substance to adsorb to XAD. In general, extraction efficiencies tend to be higher in salt water. The ionic strength of surface water is generally too low to have an effect on adsorption coefficients. Ionic strength may play a role in sewage treatment plant effluent (with high ionic strengths), however.

*Humic acids*

Humic acids can influence extraction efficiency by occupying sorption space on XAD. Due to their molecule size humic acids adsorb to XAD-8 rather than XAD-4. Fulvic acids do adsorb to XAD-4. By doing so, they lower the adsorption capacity of the resin. They can also reduce the desorption efficiency (Aiken et al., 1979). This may have consequences for relatively polar substances and competition for adsorption space with humic acids may occur (Struijs and Van de Kamp, 2001).

The binding of humic acids is reversible. However, as humic substances do not dissolve in organic solvents, they will not be eluted with acetone (Struijs and Van Buren, 1995). As a result, humic substances will probably not be present in the water extract, and they do not pose a risk of unintentional toxicity in the water concentrate.

*Particulate matter*

The XAD-technique will not isolate substances that are adsorbed to particulate matter (AWWA-KIWA: 1988). See also section 6.2.3.

Because the concentration procedure should reflect the biological availability of the substances in the original sample as good as possible, it has been chosen not to adjust any of these water parameters during extraction.

**Substance properties**

In order to be extracted from water by XAD, the properties of substances must meet the following criteria (AWWA-KIWA, 1988):

- the molecules must be partly hydrophilic and partly hydrophobic;

- the affinity of the hydrophilic part of the molecule for water must be lower than the affinity of the hydrophobic part of the molecule for XAD.

The adsorption process is defined by the interaction between the XAD and the organic substances. This interaction is a combination of three parameters: physical adsorption, solution and steric exclusion. The composition of a molecule determines whether it is organic or inorganic, hydrophobic or hydrophilic, polar or non-polar, and big or small. These properties influence the affinity of the molecule for the XAD.

*Organic micropollutants*
Both volatile and less volatile substances of an apolar or weakly polar lipophilic nature are readily adsorbed to these macroporous resins. This kind of chemicals is considered to be able to pass through biological membranes, and is thus biologically active to some extent (Slooff et al., 1984). In general, substances with a log $K_{ow} \geq 2$ are expected to be extracted from water to a large extent (Struijs and Van Buren, 1995).

*Inorganic micropollutants*
XAD-4 and XAD-8 resins are not able to extract inorganic ionogenic substances such as heavy metals and inorganic salts from water. This implies also that inorganic salts do not disturb the toxicity measurements. However, it also means that the contribution of heavy metals is not included in the toxic effects measured using the water concentrates. Calculations with models can provide an indication of metal toxicity (see Appendix III).

*Hydrophobicity and hydrophilicity of substances*
Extremely hydrophobic substances tend to adsorb to particulate matter and to the surface of the extraction vessel rather than to XAD. They are not likely to be extracted from water effectively, also because of their low solubility in water. The extraction of hydrophilic substances depends on the affinity of the hydrophylic part of a molecule to water compared to the affinity of the hydrophobic part of a molecule to XAD (Collombon, 2007). Little is known about substances with log $K_{ow} \leq 2$. In practice, however, substances with Log $K_{ow}$ as low as -2.2 have also been extracted using the combination of XAD-4/8 (Van Stee et al., 2002), although extraction efficiencies are unknown. A log $K_{ow} \leq 2$ is nevertheless considered as indicative for reduced recovery because Collombon (2007) found substantial loss at the extraction stage for several substances with this property.

*Polar substances*
Polar organic substances can be extracted using XAD-4/8, provided that the molecule partly consists of a relatively apolar side with more affinity for the XAD than for water. The molecule may need extra time to ‘fit’ to the adsorption spaces and may therefore need more extraction time than substances that are more non-polar. In general, these hydrophilic substances have poor recoveries.

*Steric exclusion*
Materials for adsorption must be able to migrate through the pores of the XAD to the adsorbing surface. Large molecules and molecule clusters have low adsorption efficiency due to their ‘particle diameter’(AWWA-KIWA, 1988).

### Environmental parameters

As discussed, several parameters have a bearing on extraction, which are summarized in Table 3-2. So far, no experimental concessions have been made to reduce the influence of environmental parameters during extraction.

*Temperature*
Adsorption of hydrophobic substances increases as the temperature rises. Adsorption of hydrophilic substances such as nonionogenic surfactants decreases. Therefore we chose to carry out the extractions at ambient temperature.

*Light*
Surface water samples usually contain different forms of life, e.g. algae. In light, algae continue primary production, resulting in an increase in pH, which in turn can change the effectiveness of the adsorption of chemicals. Photodegradation of components may also occur. For this reason, extractions were carried out under reduced light conditions.

In addition to the parameters mentioned in Table 3-2, it was also checked if the concentration of organic solvents in the water could disturb the extraction. It was found that only above a solvent concentration of 20 % (!) extraction efficiency decreased. Such values are far above values found in surface waters.

**Table 3-2. Summary of the influence of some parameters on the extraction efficiency. For more detailed information see Struijs and Van Buren (1995).**

| Parameter | Substance group | extraction efficiency when parameter value increases |
|---|---|---|
| pH | acids | - |
| | bases | + |
| Ionic strength, NaCl | nitrosamines | + |
| | nonionogenic detergents | + |
| Temperature | hydrophobic | + |
| | hydrophilic | - |
| Water solubility | extremely hydrophobic | - |
| | weakly hydrophobic | + |
| | weakly hydrophilic | + |
| | extremely hydrophilic | - |

### 3.3.2 Elution from XAD

**Solvent selection**

In order to be extracted from the XAD, a substance must be desorbed from the resins. The choice of solvent, or combination of solvents, is a key factor in the effectiveness of elution. The choice of solvent for the purpose of substances concentration was based on solvent capacity, boiling point, vapour pressure, toxicity in bioassays and health safety procedures.

In principle, several solvents are suitable for elution of XAD, including liquid carbon dioxide, using supercritical fluid extraction (Struijs et al., 1998). For several reasons, acetone proved to be most suitable for the purpose (Struijs and Van de Kamp, 2001). It was selected because of its general properties as a solvent – a wide range of chemicals are soluble in acetone – its volatility and its low boiling point (56 °C), which make it easy to remove (see textbox on solvent properties). The solvent must be free of impurities to avoid addition of unintentional toxicity.

**Solvent properties**

For desorption from XAD, the substance must be soluble in the solvent that is used to elute the XAD. Preferably, the substance should have more affinity for the solvent than for XAD. The affinity can be characterised by the Hildebrand solubility parameter δ, which is related to the polarity of a substance. The higher δ, the more polar a substance. For good desorption, δ of the substance should be closer to δ of the solvent than to δ of the XAD. Generally, compatibility between two materials can be expected when their solubility parameters are close in value.

**Table 3-3. The affinity of several substances expressed as the Hildebrand solubility parameter δ (From Burke, 1984).**

| Substance | δ ($J^{1/2}/m^{3/2}$) |
|---|---|
| Methanol | 29.7 |
| Ethanol | 26.2 |
| DMSO | 26.4 |
| Water | 23.5 |
| Dichloromethane | 20.2 |
| Acetone | 19.7 |
| XAD-8 | 19.3 |
| XAD-4 | 18.6 |
| Hexane | 14.9 |

Acetone concentrates (25 ml) are stored at -20 °C to minimise loss of substances. Samples may be stored for several months before analysis.

**Methodical influences**

The duration of the drying of XAD, the acetone volume used for elution and the flow rate during elution can influence recovery of substances.

*Drying of XAD*
XAD must be dried to get rid of any excess water that will affect distillation. In general, compounds that dissolve well in acetone will come off the XAD after drying overnight. In spite of drying, however, some water will remain in the XAD pores. Acetone mixes well with water, so the remaining water can be eluted from the XAD with acetone. The acetone eluate will thus always contain a certain amount of water.

Some of the more polar or hydrophilic substances, particularly substances that do not dissolve well in acetone, may be eluted with the remaining water in the XAD pores. The drier the XAD, the poorer the recovery of this kind of component. On the other hand, too much water in an acetone eluate may lead to incomplete distillation (section 3.3). Because acetone can lead to toxicity, drying time was optimized for distillation efficiency.

The amount of water can be checked by weighing the XAD before and after drying. Dried XAD should not weigh more than 0.3 g per ml. When using 2.5 ml XAD for a sample of 10 litres, a dried XAD sample should not weigh more than 0.75 g.

*Acetone elution*
The volume of acetone used to desorb the chemicals from the XAD should be large enough to elute them to a high extent, and small enough to keep the time needed to evaporate or distil the acetone from the sample to a minimum. The acetone volume has therefore been standardized at about 1.5 times the XAD wet volume.

*Elution column shape*
The elution column should be narrow, to ensure contact between XAD and acetone is efficient. The height to diameter ratio of the XAD in the elution column should be no less than 3:1.

*Acetone flow rate*
In order to stimulate desorption of the chemicals, the flow rate should be kept low during acetone elution, at approximately 1 ml/min.

### 3.3.3 Kuderna-Danish distillation

For toxicity testing, a sample needs to be free of any substances related to the treatment of the sample that can cause unintended toxic effects. An acetone eluate is not suitable for toxicity testing, so the acetone must be removed. Kuderna-Danish distillation (KD-distillation) is employed to remove the majority of the acetone from a sample. This type of distillation is widely used for concentration of chemicals by evaporating the solvent.
Care must be taken that the XAD does not contain too much residual water, as boiling of the acetone eluate will cease before the residue volume is sufficiently reduced.
An acetone eluate that has been dried overnight usually contains little water. On the other hand, when the amount of water is too small, the sample may boil dry during distillation. This will result in complete loss due to evaporation and precipitation of chemicals. A controlled amount of water (0.5 ml) is therefore added to the acetone eluate at the start of the distillation to prevent it from boiling dry. The distillation residue (0.2 ml) generally consists of approx. 40 % water, the remainder is acetone. If a distillation residue volume is too large, the water concentrate will contain too much acetone and will cause background toxicity.

Water concentrates need to be analysed as soon as possible after conversion, preferably within 24-48 hours. If the test laboratory cannot process the samples immediately for bio-analysis, the water samples must be stored at –20 °C. ISO 5667-3 (2003) prescribes a storage time not exceeding two weeks.

**Methodical influences**

During Kuderna-Danish distillation at 65-70 °C, acetone is removed from the sample and the volume of the sample is drastically reduced. Volatility, stability, heat resistance and water solubility of substances determine whether they will be recovered in the water concentrates.

*Volatility of substances*

Acetone volatility makes it easy to remove from the samples. Although even volatile chemicals have been recovered after KD-distillation (Struijs and Van de Kamp, 2001), some of them may be lost during distillation. Keeping the evaporation/distillation time to a minimum will also minimise the loss of volatile chemicals. Towards the end of distillation, the acetone-water mixture becomes an azeotropic mixture, so not all the acetone can be removed.

*Stability/heat resistance of substances*

Besides evaporation, decomposition of chemicals, or reaction of chemicals with each other may occur at the elevated temperature during distillation. Distillation duration can therefore influence the recovery of substances that decompose, react with other chemicals or are not resistant to elevated temperatures in any other way.

*Water solubility of substances*

Concentrations of many hydrophobic organic pollutants in surface water are in the range of ng/l - µg/l. After a theoretical thousand fold concentration, these concentrations range from µg/l – mg/l. Water solubility of chemicals that are in the order of mg/l should therefore be high enough to be kept in solution. The combination of chemicals concentrated in a small volume of water may result in a sum of chemicals that exceeds the maximum solubility of some of the chemicals. Adsorption of (extremely) hydrophobic chemicals to glass may be the result. Due to this loss of some chemicals may occur. However, after KD-distillation some acetone remains in the water concentrate, which will act as a co-solvent. As a result of this, the solubility of chemicals in the water concentrate will be somewhat elevated.

## 3.4 Performance of the concentration procedure

Several methodological experiments have led to the development of an optimized concentration procedure. The details of the standard concentration procedure are summarized in Table 3-4. Chemical recovery of different mixtures of chemicals was tested with this procedure (Struijs and Van de Kamp, 2001). Besides recovery of substances, also some methodical factors are examined, like:

- influence of the presence of humic acids;
- the extraction duration and XAD:water volume ratio;
- the influence of drying XAD;
- the influence of storage time of concentrated samples.

The chemicals used were selected for a combination of properties: hydrophobicity, volatility, toxicological relevance (occurrence in the environment), mode of action (non-specific narcotic toxicity, specific toxicity), and potential for chemical analysis. Additionally, Maas and Van den Heuvel-Greve

(2005) have investigated the recovery percentages of WFD priority pollutants with the standard method.

**Table 3-4. Details of the standard concentration procedure.**

| Method specifications | Standard chemical validation |
|---|---|
| Sample volume (l) | 60 l. |
| XAD-4 volume (in water) (ml) | 7.5 |
| XAD-8 volume (in water) (ml) | 7.5 |
| Drying of XAD (h) | approx. 18 |
| Acetone elution volume (ml) | 25 |
| KD-distillation temperature (°C) | 65-70 |
| KD-distillation addition of water (ml) | 0.5 |

The following test mixtures of chemicals with different properties were tested:
- Narcotic mixture: a mixture with hydrophobic chemicals with log Kow ranging from 2.6 to 4.8, concentrations ranging from 0.3-45 µg-l and including some relatively volatile chemicals;
- Pesticide mixture: 11 pesticides were selected, including some very persistent (lindane) and some less persistent pesticides. Due to different chemical analyses, the pesticides were divided into two mixtures that were tested separately. However, the results are presented as one group of chemicals;
- Tributyltin and Triphenyltin (TBT/TPT);
- Three surfactants: surfactants are high production volume chemicals (HPVs). The selected surfactants are considered to be representative of anionic (LAS and dodecylsulphonate) and nonionic surfactants (octaethylene glycol monotetradecyl ether) (Struijs and Van de Kamp, 2001);
- WFD-mixture: all the priority pollutants of the WFD, with the exception of metals.

The first four mixtures were spiked into mineral water (Spa blue) at low concentrations (µg/l). In order to measure the influence of humic acids the same mixtures were made in mineral water enriched with 10 mg/l humic acid and in natural surface water from the Amsterdam-Rhine canal (ARK). The priority pollutants of the WFD were spiked to standard medium (Dutch Standard Water) in concentrations based on the level of the Environmental Quality Standards (EQS; Maas and Van den Heuvel-Greve, 2005).

The recovery of various chemicals and the influence of some methodological factors are evaluated in section 3.4.1. to 3.4.5. This evaluation is followed by sections on repeatability (3.4.6), conclusions (3.4.7) and recommendations (3.4.8).

### 3.4.1 Recovery of various chemicals

The concentration of chemicals was measured in the water sample after extraction, in the acetone concentrates and in the water concentrates after distillation. The concentrations of the priority pollutants of the WFD were only measured after spiking to standard medium (initial concentration) and in the water concentrates after distillation (Maas and Van den Heuvel-Greve, 2005).

**Narcotic mixture**

The extraction of the narcotic mixture from water was good and was in the range of 88 – 100 % (Figure 3-5). The recovery was slightly declined after elution with acetone, but was still 88 % in average. The recovery in the water concentrates was very variable for the narcotic mixture. The average recovery of the substances was 58 % ± S.D. 22 %. Volatility of a substance seemed to be a major

factor. Substances with a vapour pressure of 28 Pa or more at 20-25 °C showed recoveries below 50 % of the nominal initial concentration. All other substances that were examined were relatively non-volatile, and loss of these chemicals in this part of the procedure was limited (Struijs and Van de Kamp, 2001).

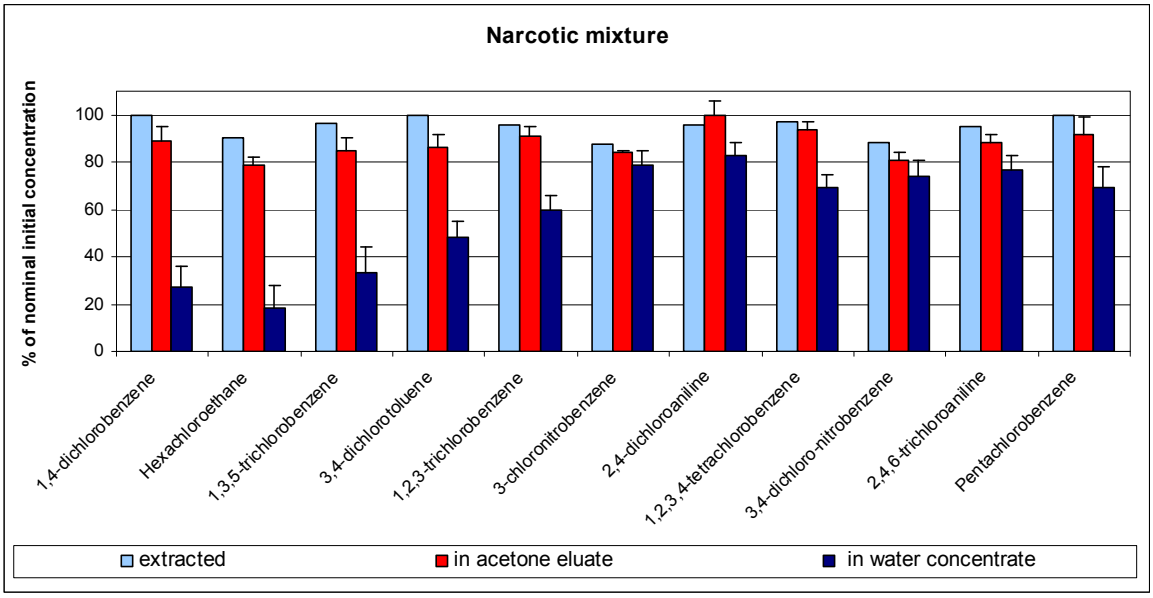


**Figure 3-5. Average recoveries of chemicals (% of nominal initial concentrations, with standard deviation) in the narcotic mixture after extraction with XAD, after elution of XAD with acetone and in water concentrates (Struijs and Van de Kamp, 2001).**

### Pesticide mixture

The extraction of pesticides from the water showed good recoveries (> 90 %) with the exception of mevinphos (63 %) and metoxuron (74 %) (Figure 3-6). The latter two substances have a log $K_{ow}$ of 1.2 and 1.6, respectively, which means a relatively high affinity for water. Extraction may therefore be less efficient. The recovery of pesticides after elution with acetone was also slightly lower than the recoveries of the water extraction. The average recovery in acetone was 86 % ± 13 % of the nominal initial concentrations of the pesticide mixture.
Pesticide recovery in water concentrates was 70 ± 10 % average. Triazinphos (81 ± 7 %) and chlorfenvinphos (81 ± 3 %) showed the best recoveries for the pesticides. The mevinphos recovery was the lowest (50 ± 3 %) (Struijs and Van de Kamp, 2001).

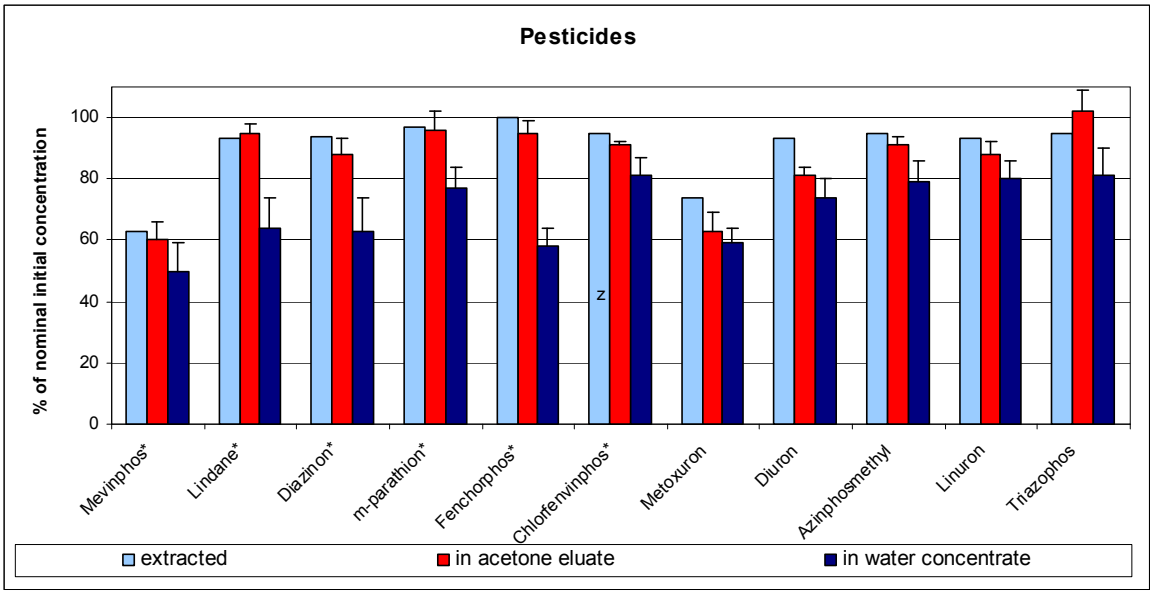


**Figure 3-6. Recoveries of chemicals (% of nominal initial concentrations, with standard deviation) in the pesticide mixtures after extraction with XAD, after elution of XAD with acetone and in water concentrates (Struijs and Van de Kamp, 2001). Substances marked with * were analysed with gas-chromatography, other substances were analysed by means of HPLC.**

### Organotin substances

Organotin substances were extracted from the water (87 %) as efficiently as the narcotic mixture and the pesticides (Figure 3-7a, Struijs and Van de Kamp, 2001). After elution with acetone and distillation the recovery was very low (both around 1 – 2 %).

### Detergents

Detergents react variable (Figure 3-7b, Struijs and Van de Kamp, 2001). Not of all fractions results on detergents were available due to analytical difficulties in water samples. Recoveries in the water concentrates ranged from 3 ± 0.7 % for sodium 1-dodecane sulfonate to 38 ± 7.8 % for the anionic sodium dodecanebenzene sulfonate (LAS) and 104 ± 1.4 % for the non-ionogenic octa-ethylene glycol monotetradecyl ether. This trend is in accordance to the decreasing polarity of the chosen surfactants.

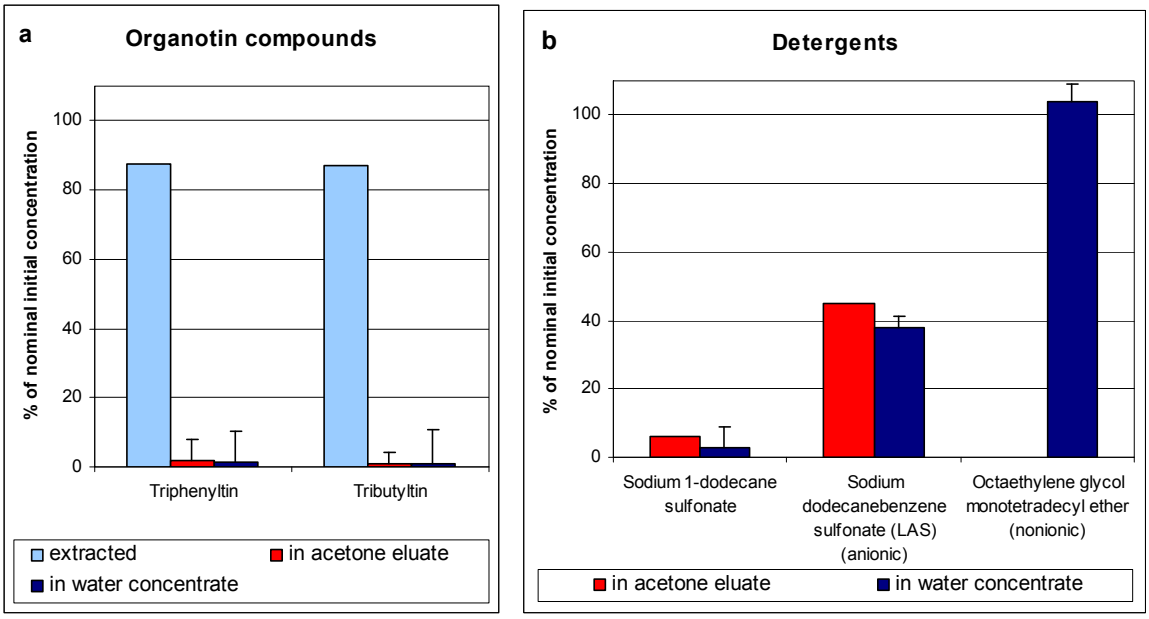


**Figure 3-7. Recoveries of organotin substances after extraction with XAD (a); and recovery of organotin substances (a) and detergents (b) after acetone elution of XAD and in water concentrates (% of nominal initial concentrations, with standard deviation). Struijs and Van de Kamp, 2001.**

### WFD priority pollutants

The recoveries for the other organic substances were fair but varied substantially between 42.5 – >100 % (values > 100 % are measurement artefacts, for results see Figure 3-8). During the whole procedure volatile substances (VOC) almost disappeared with highest recovery for 1,2,4-trichlorobenzene (28.5 %). Recoveries of the volatile PAHs (antracene; 17.9 % and naphthalene; 3.4 %) were also poor. However, also surfactants and DEHP (15.8 %) showed a poor recovery.

### Hydrophylic chemicals

Recently, questions were raised about the suitability of the method for new, 'upcoming' chemicals, i.e. chemicals that only in the last decade have been recognized as a potential environmental problem. These chemicals are of interest because of their occurrence in the aquatic environment and their (unknown) potential ecological impact. In Collombon (2007) information on suitability of the XAD-concentration method was collected and the recovery of some of these chemicals has been investigated with spiked mineral water 'Spa blauw' (Figure 3-9).

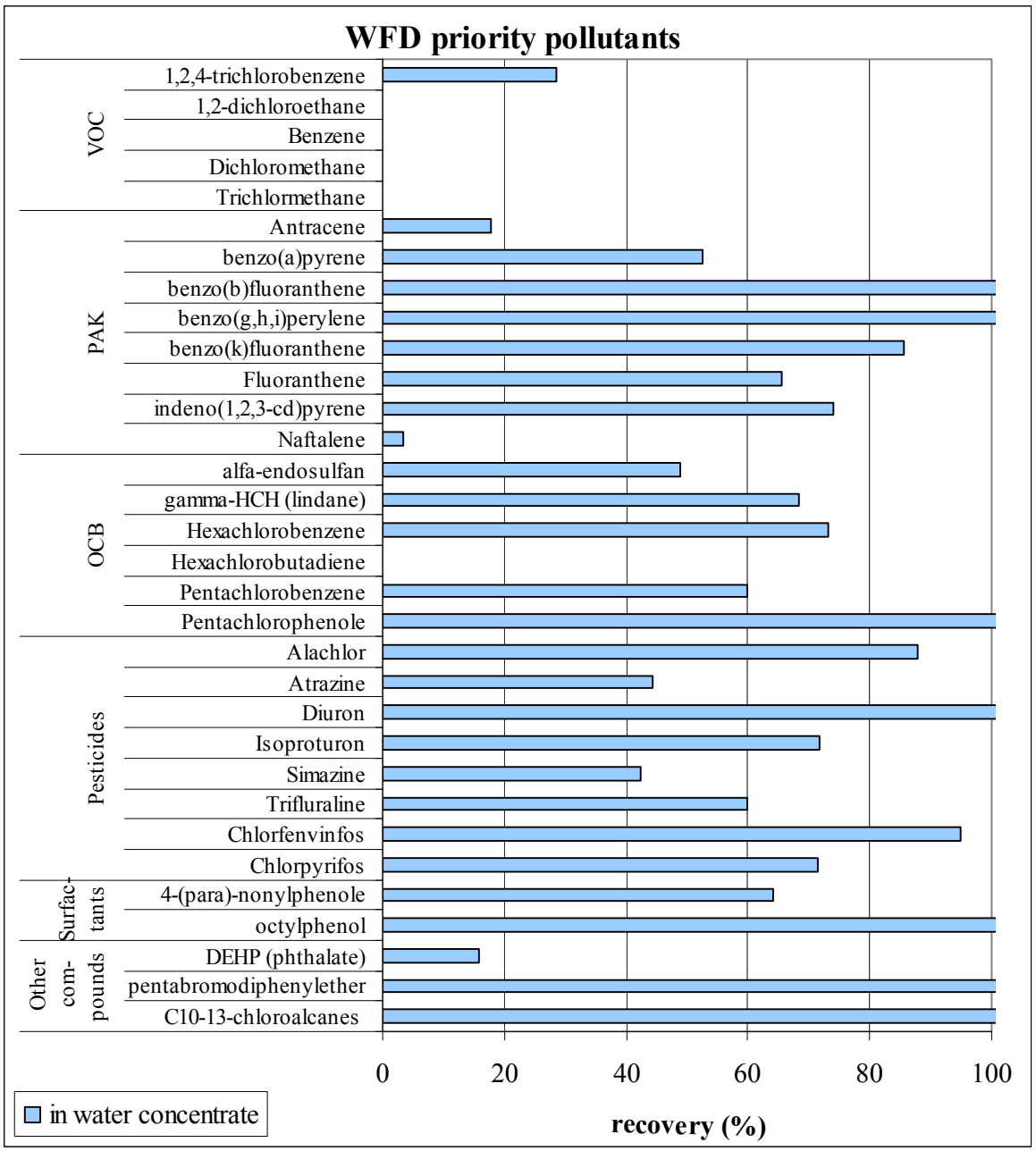


**Figure 3-8. The extraction efficiency of WFD priority pollutants as percentage of the actual initial concentration (Maas and Van den Heuvel-Greve, 2005). All substances were spiked in medium, but some disappeared during the procedure.**

The substances, which were selected on physical-chemical properties (log $K_{ow}$ < 2), specific toxic mode of action and chemical analysis possibilities, show extraction recoveries from 1 to 58 % of the measured initial concentration, with an average of 30 %. No correlation was found between log $K_{ow}$ and extraction efficiency.

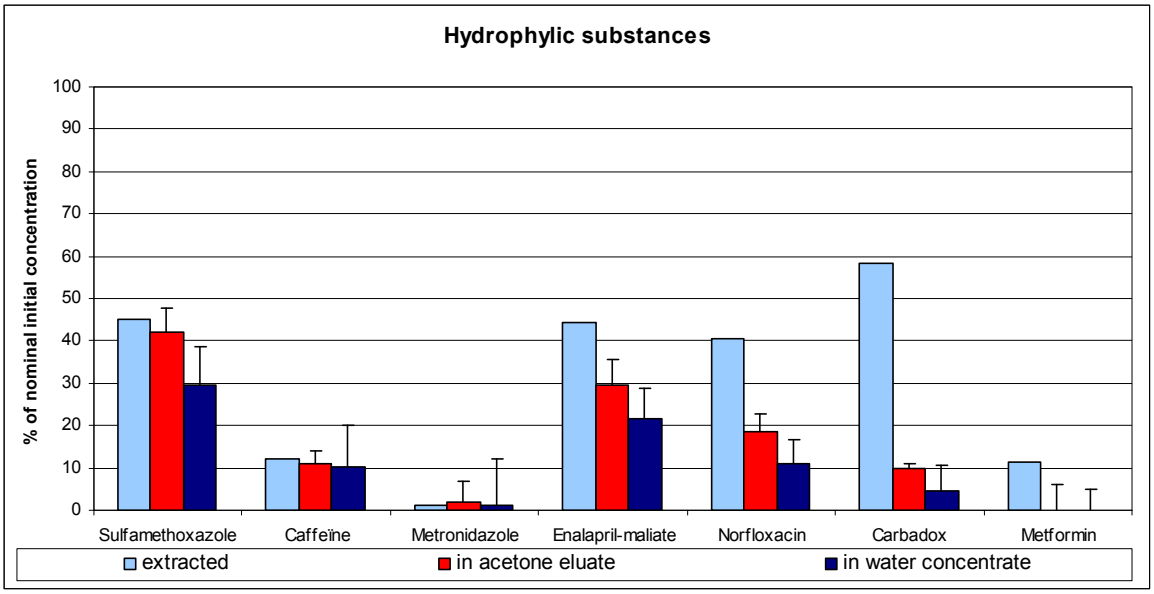


**Figure 3-9. The extraction efficiency for a selection of hydrophilic substances in spiked mineral water ('Spa blauw') from Collombon, 2007.**

Elution recoveries for these substances show an obvious correlation with log $K_{ow}$, with a high elution recovery for sulfamethoxazole (log $K_{ow}$ 0.89) and methformin (log $K_{ow}$ -2.64) completely lost. After distillation and replenishing of the residues, the average overall recovery percentage in the water was 11 %. In the water concentrates an average of 31 % was lost with respect to the acetone eluates. All recoveries from this study should be interpreted as indicative, as the analytical method was not validated yet (Collombon, 2007).

### Conclusions for recovery

The chemical analysis results of the substance mixtures indicate that extraction and elution efficiencies are generally good for hydrophobic substances. More hydrophilic substances show substantial loss at the extraction stage. A log $K_{ow} < 2$ may be considered as indicative for reduced recovery in the concentration procedure. The elutions of substances show an obvious correlation with log $K_{ow}$. For recovery after Kuderna-Danish distillation, volatility of a substance is a major factor. Substances with a vapour pressure of 28 Pa or more at 20-25 °C showed recoveries below 50 % of the nominal initial concentration. The overall recovery for substances, calculated from the (non-volatile) WFD priority pollutants is: 77 ± 26 %.

## 3.4.2 Influence of humic acid

The influence of the presence of humic substances on substance concentration efficiency was investigated for the narcotic mixture, the pesticide mixture and the detergents. These mixtures were tested in pure mineral water, mineral water with of humic acids or natural surface water (Struijs and Van de Kamp, 2001). An Effect of humic acids on recovery was examined only in acetone eluates and water concentrates. Concentrations in spiked water and waste water were not measured.

In almost all of the cases the extraction and elution efficiency did not differ significantly for the three water types. Differences in efficiencies for individual substances in mineral water, compared to natural water with humic acids present, varied from 5 % decrease to 13 % increase (Figure 3-10).

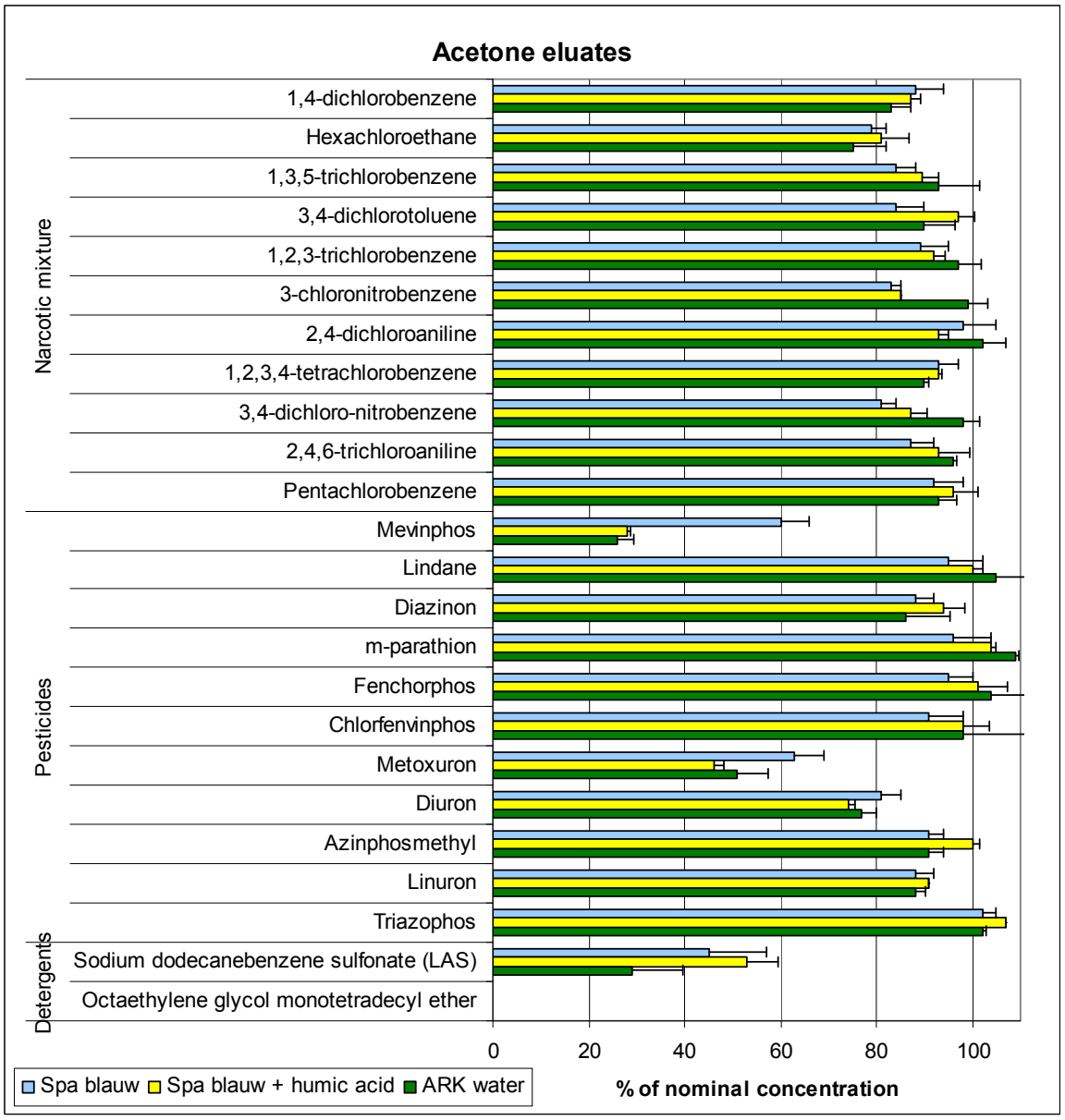


**Figure 3-10. Average recovery (±S.D.) in acetone eluates after XAD-extraction with spiked mineral water ('Spa blauw'), spiked mineral water with humic acids (10 mg/l) added, and spiked natural surface water (from the Amsterdam-Rhine canal).**

Exceptions are mevinphos and metoxuron in the pesticides mixture. These pesticides show significantly lower recoveries in the acetone eluate of mineral water samples with humic acid and natural water samples than the mineral water samples. Recovery of mevinphos in mineral water with humic acids shows a recovery of 28 % of the nominal initial concentration, and 26 % of the spiked concentration in natural surface water, which is lower than recovery from untreated mineral water (60%). Metoxuron has a somewhat higher recovery of 46 % (mineral water + humic acid) and 51 % (natural surface water) compared to 63 % (mineral water). This indicates that there may be competition for adsorption sites between the substances and humic acids present in the water at this amount of XAD.

After KD-distillation the recovery of the narcotic and pesticide mixture was more or less the same for the three water types (Figure 3-11). The results of the procedure for the three treatments indicate that the presence of humic substances is not of great influence on the recovery for most of the narcotic and pesticide substances.

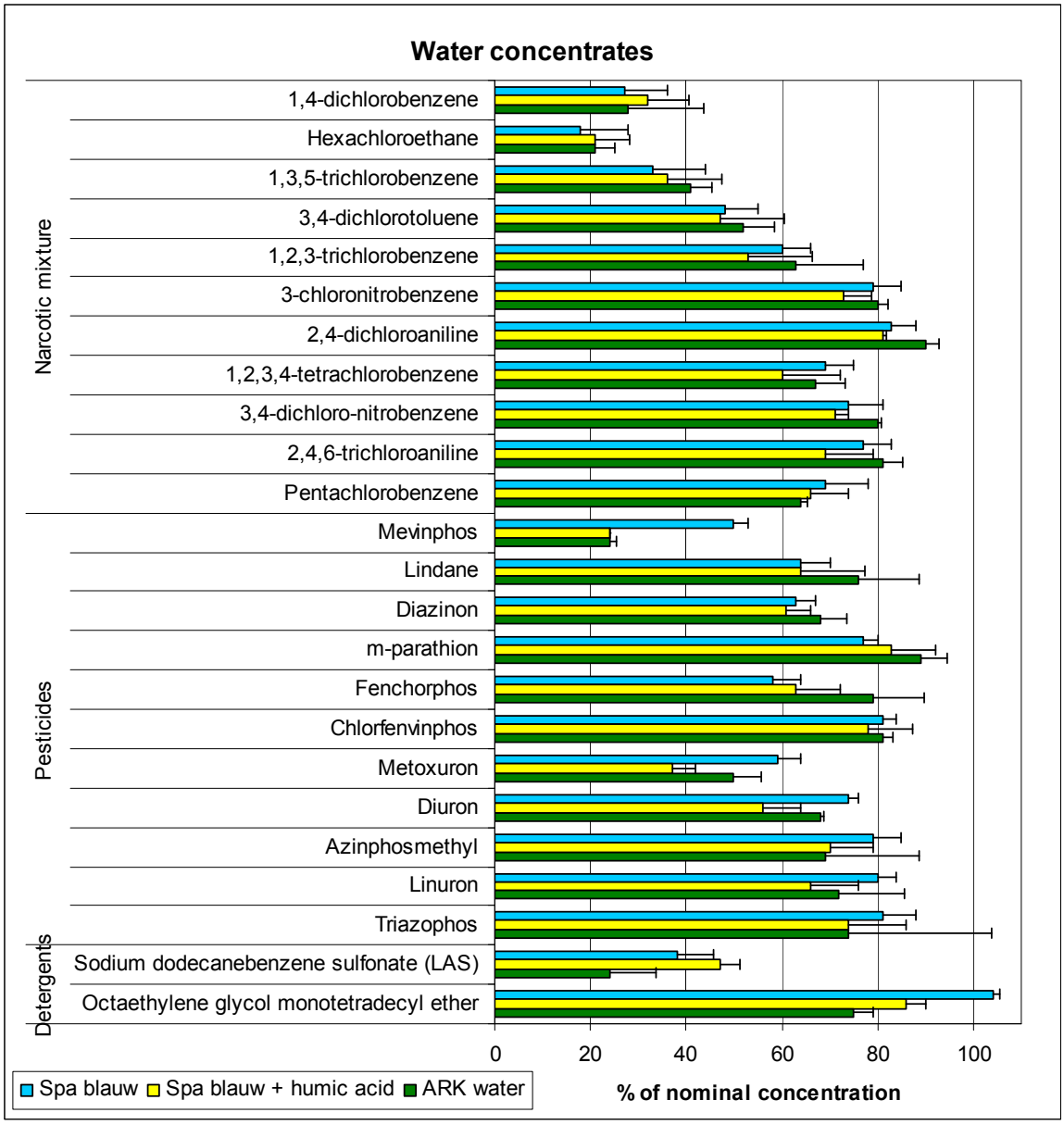


**Figure 3-11. Average recovery (±S.D.) in water concentrates after XAD-extraction with spiked mineral water ('Spa blauw'), spiked mineral water with humic acids (10 mg/l) added, and spiked natural surface water (from the Amsterdam-Rhine Canal).**

### 3.4.3 Influence of XAD/ water volume and extraction duration

Struijs and Van de Kamp (2001) have investigated the influence of different XAD/ water volume ratios and the extraction duration on the depletion of substances from water. The motive to investigate these relations was to develop a less laborious and material consuming method. To be able to handle the

sample with a single KD-distillation, the amount of acetone had to be reduced. Therefore, the amount of XAD should be reduced for about a factor 10 compared to the method which was in uses at that moment, which reduced the costs of the method as well.

The recoveries of the different substances in the narcotic mixture were investigated for 1, 2, 2.5, 10, 20 and 50 ml XAD/10 l and different extraction periods. Although for some substances the extraction was obviously less at lower XAD amounts, the average decrease in recovery was acceptable taking the practical and financial advantages into account (Figure 3-12). To reduce the disadvantages of this reduction of XAD amount, the extraction time was increased from 24 to 48 hours. Also, the acetone bed volume (volume of acetone used compared to the XAD volume) increased from 1 to about 1.5.

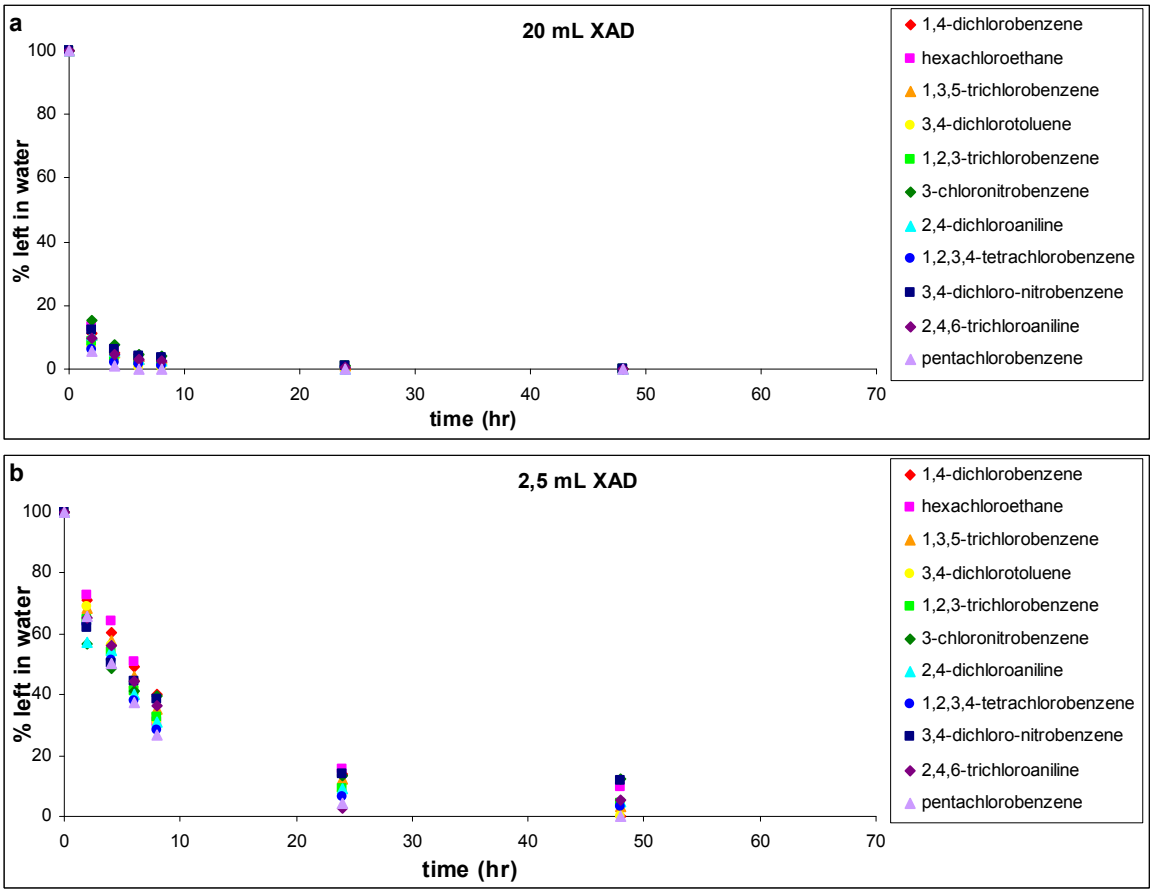


**Figure 3-12. Depletion of 11 hydrophobic chemicals (narcotic mixture) from a water sample as they adsorb onto XAD resins with 20ml XAD/ 10 l (a) and 2,5 ml XAD/ 10 l.**

### Influence of drying XAD

After extraction with XAD the resin is sieved (50 µm mesh) and transferred to a Petri dish. The Petri dish is placed during the night (±18 hours) in a gentle air stream in a hood in order to dry the XAD. Any excess of water that will affect distillation must be eliminated from the XAD. Dried XAD should not weigh more than 0.3 g/ml, before it is be eluted with acetone. Subsequently the acetone will be removed by distillation during 20 – 30 minutes. Boiling of the mixture ceased when an azeotropic mixture of acetone and water was formed. This residue has a volume of approximately 0.2 ml.

Too much water in an acetone eluate may lead to distillation problems, as boiling of the acetone eluate will cease before the residue volume is sufficiently reduced. A relatively high volume of residue will contain an acetone concentration that exceeds the no-effect concentration level for some test organisms.

Struijs and Van de Kamp (2001) found an average concentration of acetone of 0.19 volume percent with a 95 percentile of 0.29 volume percent in thousand fold concentrated water samples. The no-effect concentration of acetone derived by Vaal and Folkerts (1998) is however 0.15 volume percent. When data for other test species are considered, an acetone concentration of roughly 2 to 3 volume % seems to be the lowest concentration where effects can be expected (Maas, unpublished data). Using an estimated value of at most 2.5 volume % of acetone (Collombon, unpublished data), Probably none of the test organisms will be seriously affected, perhaps except for organisms exposed to the highest concentration factor.

### 3.4.4 Influence of storage time

After extraction with XAD, samples will be eluted with acetone as soon as possible. Storing dry XAD at 4 °C in an excicator does not alter the recovery of the narcotic mixture significantly and could be an alternative for storing acetone eluates. The recovery of hexachloroethane, 1,3,5- and 1,2,3-trichlorobenzene, 1,2,3,4-tetrachlorobenzene, pentachlorobenzene, lindane and trichlorphos did not alter significantly after storing dry XAD (Struijs and Van de Kamp, 2001). However, this is tested only for this mixture; behaviour of other substances under these conditions was not examined. Therefore, elution of XAD with acetone should be done within the time limits that are set.

Until bioassays are being carried out the samples are stored as acetone eluates with a volume of approximately 25 ml at -20 °C. All biological tests have to be performed in a short period, because the whole acetone eluate is concentrated to yield only one batch of water concentrate.

Struijs and Van de Kamp (2001) also monitored the loss of chemicals from concentrated water samples. In these water concentrates, stored at 4 °C in glass vessels, the most hydrophobic chemicals ($K_{ow} > 4$) lost 20 – 30 % of their initial concentration after one week, most likely due to sorption on the glass wall.

### 3.4.5 Additional remarks

#### Mass balances and recovery

Results of recovery measurements in this report are reported as percentage of nominal initial concentration, i.e. the concentration that was aimed to be dissolved in the water. For the narcotic test mixture, it is known that the measured concentrations (or actual initial concentrations) are generally high, average 92 % of the nominal concentration (Table 3-5).

**Table 3-5. Average measured concentration as percentage of actual initial concentrations of the narcotic mixture for 2 datasets.**

| | Dataset 1 | | Dataset 2 | |
|---|---|---|---|---|
| | Average % (n=10) | S.D. | Average % (n=3) | S.D. |
| 1,4-dichlorobenzene | 98 | 2 | 100 | 2 |
| Hexachloroethane | 90 | 3 | 99 | 1 |
| 1,3,5-trichlorobenzene | 92 | 3 | 97 | 2 |
| 3,4-dichlorotoluene | 100 | 3 | 99 | 2 |
| 1,2,3-trichlorobenzene | 93 | 2 | 102 | 1 |
| 3-chloronitrobenzene | 86 | 4 | 95 | 1 |
| 2,4-dichloroaniline | 84 | 3 | 76 | 1 |
| 1,2,3,4-tetrachlorobenzene | 96 | 3 | 101 | 2 |
| 3,4-dichloro-nitrobenzene | 86 | 5 | 95 | 1 |
| 2,4,6-trichloroaniline | 96 | 6 | 97 | 1 |
| Pentachlorobenzene | 96 | 4 | 99 | 1 |
| Average | 92 | 5 | 96 | 7 |

However, not much is known about the actual initial concentrations of the other chemicals that were tested. Artefacts in chemical analysis may not be noticed, but also interpretation of chemical recoveries can easily be different when the actual concentration differs from the nominal concentration. Therefore it is not possible to draw 'hard' conclusions on the effectiveness of the method for the pesticides, the organotin substances and the detergents, especially when the behaviour of the test substance is not known.

When data are lacking on initial concentrations in spiked water, an indication on extraction efficiencies can be obtained from chemical analysis in acetone after elution. Although this may be a legitimate way of thinking, misinterpretation of results may occur when mistakes were made, e.g. accidental elevated or lowered initial concentrations due to erroneous spiking of water.

Without data on initial concentrations in spiked water and concentrations in waste water, it is not possible to produce a mass balance. Mass balances are helpful in recognising or excluding mistakes or artefacts due to chemical analysis. When the aim is to validate or to improve the method, it is advised to chemically check all steps in the procedure when possible.

### Optimization

The narcotic mixture comprises of hydrophobic substances that were expected to give good extraction recoveries. It should be noted that the method development was based mainly on the narcotic mixture. The choices for XAD/ water volume ratio, the extraction time, the elution volume and the KD-procedure were all based on experiments only with this mixture.

For the narcotic test mixture it is therefore known, also from former experiments during method development, that the XAD-concentration method works well, as long as the chemicals are not volatile, and less satisfactory for the relatively volatile substances (Struijs and Van de Kamp, 2001; Collombon, 2007). However, it is not known to what extent the method can be optimized for chemicals that are less hydrophobic.

The experiments with other chemicals indicate that the method may not always be as effective as it is for the narcotic mixture chemicals. Substances that are less hydrophobic and more water soluble may not be extracted from water so well, e.g. the pesticides mevinphos and metoxuron, but also bentazone (see results in Collombon et al., 1997). Moreover, chemicals that were extracted efficiently from water may not come off the XAD due to their insolubility in acetone, as indicated by the results of the organotin substances.

Nevertheless, it should be kept in mind that even with a concentration efficiency of only 10 %, the substance concentration is increased a hundred-fold with the usual 1000-fold concentration, which may be enough to cause acute toxic effects, or at least add up to the effect of the combined test chemicals in the test mixture.

**Repeatability**

Although the experimental results may give a good indication for the effectiveness of the method for the chosen substances, it needs to be said that the results are based on a limited number of experiments (usually 1, sometimes 2 with the narcotic mixture) with a limited number of replicates (Table 3-6). These experiments were carried out in one laboratory.

**Table 3-6. Chemical analysis on recovery of the test mixtures, number of replicates.**

| | Narcotic mixture | | Pesticide mixture | | Surface active substances | | Organotin substances | |
|---|---|---|---|---|---|---|---|---|
| Spiked Mineral water | 3 | | - | | - | | - | |
| Waste water | 2 | | 1, 3* | | - | | 2 | |
| Acetone eluate | 12 (2*6) | | 6 | | 2, 0 | | 2 | |
| Water concentrate | 12 (2*6) | | 6 | | 2 | | 2 | |
| | H** | ARK*** | H** | ARK*** | H** | ARK*** | H** | ARK*** |
| Spiked (Mineral) water | - | - | - | - | - | - | - | - |
| Waste water | - | - | - | - | - | - | - | - |
| Acetone eluate | 2 | 2 | 2 | 2 | 2 | 2 | - | - |
| Water concentrate | 2 | 2 | 2 | 2 | 2 | 2 | - | - |

*Pesticides were measured using different chemical analysis methods. One series of pesticides was analysed in one replicate, the other in 3 replicates.

**Spiked mineral water with 10 mg/l humic acids added.

***Spiked surface water (Amsterdam-Rhine canal)

## 3.5 Recommendations

When a method is validated, the performance of the method is known within certain ranges. The question is to what degree the concentration method should be validated. Validation to the same precision level as chemical analysis is virtually impossible, due to the lack of analytical methods and the costs of all the chemical analyses. It may not even be appropriate, because the aim of the method is to reveal toxicity of a mixture of chemicals with unknown composition.

However, to increase the logistic flexibility more attention could be paid to recovery of chemicals influenced by:

- extraction duration on recovery and eventually the relation with XAD/ water volume ratio;
- drying duration on recovery;

- acetone elution volume, to examine within which ranges the recovery is constant and within acceptable ranges;
- storage time of an acetone eluate;
- KD-distillation duration and temperature, related to initial acetone eluate volume;
- repeatability of the different steps in the concentration procedure.

Information on these subjects is needed in order to be able to judge when samples can be treated with some flexibility, and when guidelines need to be formulated stringently. This helps also to define the boundaries for certain chemical properties where the method is effective.

Specific attention can be paid to linearity of extraction efficiencies. Whether varying nominal initial concentration levels of substances may have influence on the recovery percentages of chemicals is not known. The limited number of spiked samples examined does not provide information on concentration ranges. No attention was paid yet on the influence of combining different types of chemicals or different water types (i.e. marine waters) on the recoveries.

For further validation and optimization, attention could be paid to both chemicals from the test mixtures that were used in this research, and to the behaviour of the more 'difficult' chemicals, chemicals that do not necessarily have high extraction efficiencies, or that may degrade or otherwise alter during the concentration procedure. It is also important to notice that some chemical types, like small ionogenic molecules, or extremely large molecules, probably will never give good results with this concentration method. Further research should help to define property ranges for which the method is effective. When specific groups of substances are targeted, research could be done on alternative concentration methods for these specific groups: e.g. other types of XAD, passive sampling or any other concentration method.

Since most substances are not fully recovered, it is credible that the toxicity of the concentrated sample gives an underestimate of the toxicity of the original sample. To what extent a substance contributes to the toxicity of the sample will depend for an important part on its recovery in the concentrated sample. The recovery will vary for each of the (unknown) substances in a sample, but since the chemical composition of the sample is not known, the contribution of the several substances to the toxicity is unknown. Nevertheless, it is suggested to correct for recovery in interpreting toxicity results because the concentrated sample will always give an underestimate of toxicity to some extent. Information on recovery will always be limited because for only a few substances recovery is investigated compared to the possible substances that can be found in water. Research on recovery emphasized strongly on narcotic substances in the first place. The most varied set of substances tested on recovery are the WFD priority substances (section 3.4.1), which can be concentrated except some volatile substances and hexachlorobutadiene. The average recovery for non-volatile WFD substances is 77 % .With a correction factor one should always keep in mind that toxicity found will probably not or very limited be caused by substances that are not or to a very limited extent concentrated from the original sample. In practice, a correction factor of 0.5 is used as an extra safety factor for substances which are hardly concentrated.

Furthermore it should be realized, that not all substances may be able to cause acute toxic effects after a 1000-fold concentration. Some chemicals are present in surface water in the order of ng/l, but are only acutely toxic in the range of mg/l (Derksen et al., 2001). When concentrating these chemicals 1000-fold, the maximum concentration level that can be reached is in the range of µg/l. These substances do not create an acute toxicological problem in water, but the range at which chronic toxicity occurs is likely to be lower.

## 3.6 Conclusions

The method seems suitable for concentrating chemicals that are hydrophobic and not too volatile. Less hydrophobic substances, e.g. log Kow<2 show lower concentration efficiencies. Especially when humic acids are available that can compete for sorption places, less hydrophobic substances may show lower concentration efficiencies. A higher amount of XAD could improve recovery of substances, but this has practical disadvantages (distillation in portions).

So far, a lot of effort has been put into researching the effectiveness of the method. Substances were selected, methods for chemical analysis were developed and a lot of experiments were carried out. These data can, and have to be used when validating the method. However, with the available data it is not possible to evaluate the variability of the method due to variations in sample treatments.

For this concentration method, most information that is available indicates the suitability of the method for the purpose. Because the field situation ca never be copied fully, it is impossible to give a general view on the suitability of the method.

# 4 Toxicity testing

## 4.1 Choice of test organism

The toxicity of the eluate can be measured with different bioassays. Individual test species differ considerably in their sensitivity to the bioavailable substances. To obtain the greatest possible range of sensitivity, a battery of tests should be used. Such a battery is composed of tests with organisms representing different functional groups. The choice of test organisms is limited by the small test volume of sample that is available after the concentration step. For that reason, conventional aquatic bioassays, such as a fish test, are impractical for routine environmental screening. Test duration and ethical objections are additional arguments not to add a fish test to the test battery.
This chapter gives information about the acute bioassays that have already been used on samples obtained from the XAD procedure. The choice of bioassays depends on the specific problem being addressed.

## 4.2 Selected bioassays

The selected bioassays represent three important trophic levels of an aquatic ecosystem. Tests with bacteria, algae and different invertebrates have been selected.
Willemsen et al. (1995) evaluated the use of about 30 aquatic microbiotests for general toxicity, teratogenicity and genotoxicity. Selection criteria for the composition of a test battery were:

- acute tests with whole organisms, in which the type of toxicity leads to reduced metabolism and reproduction, and/or induction of mortality;
- tests that can be performed in a small test volume;
- tests that are preferably available as a kit;
- tests that do not require specifically trained personnel or extensive laboratory facilities

In the selection procedure, the methods were compared in terms of documentation, reproducibility, sensitivity, exposure time, standardization, technical simplicity and costs. As a result, Willemsen et al. proposed a test battery using bacteria (Microtox®), algae (microplate assay), crustaceans (Thamnotoxkit F™) and preferably a Daphnia (IQ) test.

Since 1995 most practical experience in the Netherlands has been gained with these microbiotests and an additional test with a rotifer (Rotoxkit F™). This latter test was introduced to complete a battery of five tests that can be used in statistical extrapolation techniques (see chapter 5). The algae test has been replaced by a much faster alternative algae test (algae PAM). Other test organisms or promising microbiotests might also be used.

Different batteries of acute tests might be chosen depending on the application of the method. Routine monitoring programmes should preferably be performed with fast, sensitive, low-cost tests. Five or more test organisms should be used for a refined risk assessment. A test battery of three tests is sufficient to assess the effects of whole effluents. In identification studies, it might be preferable to use tests that can be compared to literature data on the effects of substances.

Table 4-1 gives an overview of the protocols used and shows the literature references that form the bases for the methods used. Table 4-2 shows the test volume needed to complete the assays and presents information about exposure time, effect parameters and endpoints.

**Table 4-1. Selected bioassays that can be performed on water concentrates.**

| Bioassay | Appendix | Protocol | References | Modifications |
|---|---|---|---|---|
| Microtox® | I.3 | RWS-RIZA | ISO 11348-3 | none |
| Algae (PAM) | I.4 | RIVM-LER | Genty et al. (1989)<br>Hofstraat et al. (1994) | Van Beusekom et al. (1999) |
| Algae (growth) | I.5 | RWS-RIZA | ISO 8692<br>Blaise et al. (1986) | exposure in multiwell plates |
| Daphnia IQ | I.6 | RWS-RIZA/ Aqua Survey Inc. (1997) | Janssen and Persoone (1993) | use of Daphnia culture |
| Daphnia immobility | I.7 | RWS-RIZA | ISO 6431<br>OECD 202 | exposure in test tubes (10 ml) |
| Thamnotoxkit F™ | I.8 | RWS-RIZA | Creasel (1992)<br>Centeno et al. (1995) | none |
| Rotoxkit F™ | I.9 | RWS-RIZA | Creasel (1990)<br>Snell and Persoone (1989) | *Brachionus rubens* replaced by *calicyflorus*. |
| Chydorid | I.10 | RIVM, UvA | Dekker et al.(2006) | volume of exposure |

**Table 4-1. Information concerning the exposure conditions and effect parameters. The volume of extract needed is based on a 1000x concentrated sample. The exposure volume is the volume in which a certain number of organisms are exposed.**

| Bioassay | extract needed | exposure volume | number of organisms | number of replicates | exposure time | effect parameter | endpoint (acute) |
|---|---|---|---|---|---|---|---|
| Microtox® | 2.5 ml | 0.5 ml | -- | 2 | 30 min | inhibition of biolumines-cence | $EC_{50}$ |
| Algae (PAM) | 4 – 5 ml | 3 ml | 1 x $10^9$ cells/l | 2 | 4.5 hours | photosynthesis efficiency | $EC_{50}$ |
| Algae (growth) | 3 ml | 270 µl | 6.6 x $10^9$ cells/l | -- | 72 hours | inhibition of growth | $EC_{50}$ |
| Daphnia IQ | 9 – 17 ml | 5 ml | 6 | 3 | 75 min | enzymatic inhibition measured by fluorescence | $EC_{50}$ |
| Daphnia acute | 20 ml | 5 ml | 5 | 4 | 48 hours | mortality/ inhibition | $LC_{50}/EC_{50}$ |
| Thamnotoxkit F™ | 7 ml | 1 ml | 10 | 3 | 24 hours | mortality | $LC_{50}$ |
| Rotoxkit F™ | 8 ml | 0.3 ml | 5 | 6 | 24 hours | mortality | $LC_{50}$ |
| Chydorid | 2 ml | 250 µl | 5 | 4 | 48 hours | mortality | $LC_{50}$ |

A brief description of each bioassay is given in the following sections. Detailed test protocols can be found in Appendix I.

### 4.2.1 Bacteria

Bacteria are important for nutrient cycling in an ecosystem. Many studies have been performed with the Microtox® assay, which is among other microbiotests one of the most sensitive (Willemsen et al., 1995). The bacteria used in this assay are sensitive to a wide range of substances, including metals, pesticides, industrial chemicals and crude oils (Calow, 1993; Azur environmental **www.azurenv.com** ).

#### Microtox® assay

The Microtox® bioluminescence assay (Bulich, 1979) is a standardized test which was commercialized in the early 1980s (Bulich and Isenberg; 1981; Microbics, 1992). The test organism is *Vibrio fischeri* (synonym for the *Photobacterium phosphoreum*), a luminescent marine bacterium (Figure 4-1). Luminescence in this species is a consequence of respiration. It diminishes as the bacterium is exposed to a chemical which interferes with the cell metabolism.
The bacteria are stored in lyophilized form, and can be used after 5 minutes of rehydration. The bacteria are incubated at 15 °C and exposed to a range of concentrations of the water extract in glass cuvettes of 3 ml. The luminescence of the bacteria is measured with a Microtox toxicity meter (M 500) after 5, 15 and 30 minutes of exposure.

The dose-response relation is estimated from the inhibition of luminescence of the bacteria by the different dilutions, expressed as a percentage of the control. Toxicity values expressed as the concentration at which 50 % of the organisms ($EC_{50}$) are affected are calculated using the Azur software package MicrotoxOmni™.

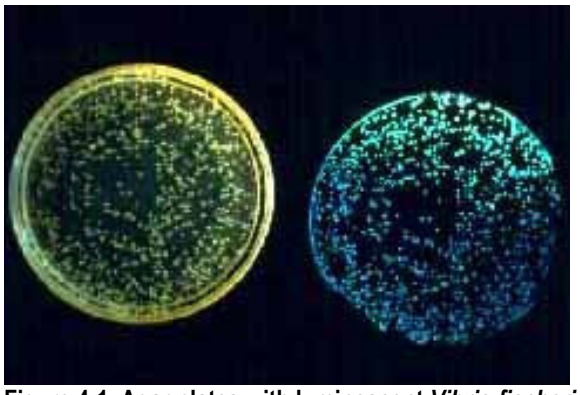

**Figure 4-1. Agar plates with luminescent *Vibrio fischeri* colonies.**

### 4.2.2 Algae

Algae are primary producers in the aquatic ecosystem and are responsible for the uptake, storage and deposition of nutrients. The growth inhibition test is sensitive to a large number of pollutants, including metals, insecticides, industrial organic components and, especially, herbicides. Inhibition of photosynthetic activity has been shown for surfactants, herbicides and oil products. In general, photosynthetic activity has been a more sensitive indicator of effect than growth.

#### Algae PAM-test

The algae PAM test with the unicellular freshwater alga *Pseudokirchneriella subcapitata* (synonym for *Raphidocelis subcapitata* and *Selenastrum capricornutum*; Figure 4-2) is an alternative to the conventional algae growth test (ISO 8692, 2004). The endpoint of this test is the inhibition of photosynthetic efficiency.

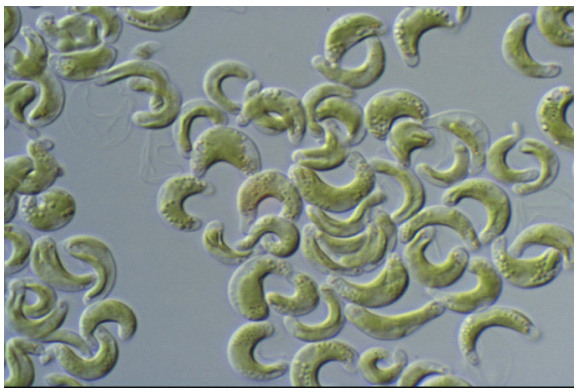

**Figure 4-2. *Pseudokirchneriella subcapitata* under the microscope.**

Algae from exponentially growing cultures are exposed to a range of dilutions of a sample. After 4.5 hours the photosynthesis of the algae is stimulated additionally with a high intensity light pulse. In the algae, the received light is converted into an energetic chemical substance which can be used to form glucose. If the energy cannot be used for photosynthesis, the energetic chemical substance will drop to its original energy level. The transmitted light intensity (fluorescence) is a measure of photosynthesis. Impairment of the photosynthesis system by toxicants will result in a change in the fluorescence level.

The Pulse-Amplitude-Modulation (PAM) fluorometer registers the changes in the fluorescence signal after stimulation with the light impulse. The photosynthesis efficiency can be calculated from the registered signals.

The dose-response relation is estimated from the photosynthesis efficiency of the different concentrations expressed as the percentage of the control. An $EC_{50}$-value is calculated using a logistic response model by Haanstra et al. (1985).

**Algae growth test**

The microplate test with the unicellular freshwater alga *Pseudokirchneriella subcapitata* is a miniaturized version of the conventional growth inhibition test (ISO 8692, 2004) that is performed in flasks. The endpoint of this test is the cell concentration measured after 72 hours of exposure.

Algae from exponentially growing cultures are incubated in polystyrene 96-well multiwell plates in continuous light in a range of concentrations of the water extract. The algal concentration is measured at the start of the incubation and after 72 hours. The cell concentration is determined by means of fluorescence measurement using a BIOTEK FL600 plate reader at 440 nm excitation and 680 nm emission. Impairment of algal growth by toxicants will result in a decrease in fluorescence.

The dose-response relation is estimated from the cell concentration of the different concentrations expressed as a percentage of the control. Concentrations at which the growth of algae was inhibited by 50 % ($EC_{50}$) were calculated using Toxcalc software (Tidepole scientific).

### 4.2.3 Invertebrates

The invertebrates used in the tests are primary consumers of algae and are food for many fish species. *Daphnia magna* is the most frequently used invertebrate in standard acute and chronic toxicity tests. Many data on the sensitivity of a wide range of substances are available in the literature. The test organisms are especially sensitive to organophosphorus insecticides.

In general, the results of the Daphnia IQ test correlate very well (r = 0.93) with the results of the conventional acute Daphnia test (Janssen et al., 1991; Zalewka-Radziwill, 2000; Aqua Survey Inc., 1997). The Daphnia IQ test is less sensitive than the conventional test to substances with a specific working mechanism such as insecticides (Vaal and Folkerts, 1998; RWS-RIZA, unpublished data).

The sensitivity of the larvae hatched from the cysts of the toxkits is comparable with other invertebrates. In general, the sensitivity of the Thamnotoxkit F™ and Rotoxkit F™ is lower than *Daphnia magna* (Vaal and Folkerts, 1998).

The Chydotox test is fairly new. It has been validated for metals (Dekker et al., 2006), and its sensitivity seems comparable to or even greater than that of *D. magna*, even to the effects of insecticides (data not published).

**Daphnia IQ test**

The Daphnia IQ test (Aqua Survey, Inc., 1997) using *Daphnia magna* (Figure 4-3) is an alternative to the conventional acute toxicity test with daphnia (ISO 6431). In the Daphnia IQ test toxicity is determined by measuring the inhibition of an enzymatic reaction in living daphnids. Daphnia can be obtained from a laboratory culture or commercially as ephippia. Hungry young water fleas of 2 to 5 days old are exposed for one hour to a range of dilutions of the water extract. After one hour they are fed with a substrate 4-methyllumbelliferyl-β-D-galactoside (MUF). Healthy, unaffected daphnia will absorb the substrate and are able to break the saccharid bond. This reaction releases fluorescent umbelliferon, which displays strong fluorescence under UV-light (Figure 4-3). The fluorescence is measured visually with the aid of a long-wave UV-lamp.

The dose-response relation is estimated from the level of fluorescence from Daphnia in the different concentrations in relation to the control. Toxicity values expressed as the concentration at which 50 % of the organisms ($EC_{50}$) are affected are calculated using Toxcalc software (Tidepole scientific).

**Acute Daphnia test**

The acute Daphnia test is a conventional standardized test (ISO 6431, 1996).Young *Daphnia magna*, less than 24 hours old at the start of the test, are exposed to a range of dilutions of the water concentrate. In contrast to the normal test procedure, the exposure volume is decreased to a minimum of 5 ml. The test duration is 48 hours. At the end of the test the number of dead or immobile animals is counted in each concentration.

The dose-response relation is estimated from the number of living, unaffected Daphnia in the concentrations in relation to the control. Toxicity values expressed as the concentration at which 50 % of the organisms are dead or immobile ($LC_{50}$/$EC_{50}$) are calculated using Toxcalc software (Tidepole scientific).

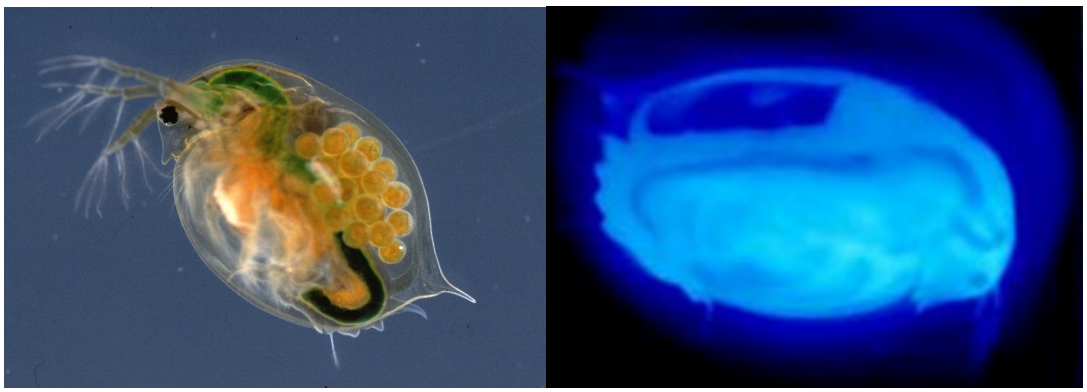

**Figure 4-3. *Daphnia magna*, adult 2 mm (left; photo made by D.Kolmeijer) and an active fluorescing *Daphnia magna* as observed in the IQ test (right).**

### Thamnotoxkit F™

Thamnotoxkit F™ is a commercially available toxkit (Creasel Ltd., Deinze, Belgium). The test organism is the crustacean *Thamnocephalus platyurus* (Centeno et al., 1995; see Figure 4-4). The endpoint of the test is lethality.

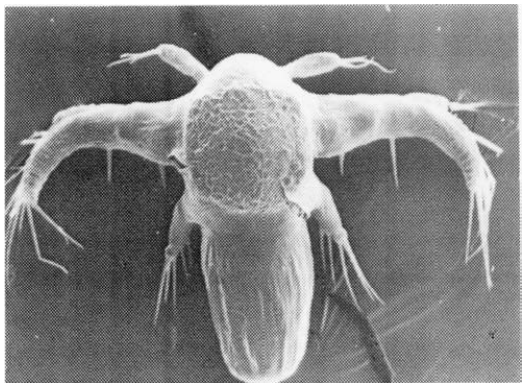

**Figure 4-4. *Thamnocephalus platyurus* under the microscope.**

The organisms are available as cysts. The test animals are hatched overnight from these cysts at 25 °C, and are ready for use after 24 hours. Exposure takes place in a multiwell microplate with a volume of 1 ml (Figure 4-5). After 24 hours of exposure, the number of dead animals is counted in each concentration.

The dose-response relation is estimated from the number of living crustaceans in the concentrations in relation to the control. Toxicity values expressed as the concentration at which 50 % of the organisms are dead ($LC_{50}$) are calculated using Toxcalc software (Tidepole scientific).

**Figure 4-5. Multiwell microplate for the performance of the test with *Thamnocephalus platyurus*.**

### Rotoxkit F™

Rotoxkit F™ is a commercially available toxkit (Creasel Ltd., Deinze, Belgium). The test organism is the rotifer *Brachionus calyciflorus* (Figure 4-6). The endpoint of the test is lethality.
The organisms are available as cysts. The test animals are hatched overnight from these cysts at 25 °C, and are ready for use after 24 hours. Exposure takes place in a multiwell microplate with a volume of 0.3 ml. After 24 hours of exposure, the number of dead animals is counted in each concentration.

The dose-response relation is estimated from the number of living rotifers in the concentrations in relation to the control. Toxicity values expressed as the concentration at which 50 % of the organisms are dead ($LC_{50}$) are calculated using Toxcalc software (Tidepole scientific).

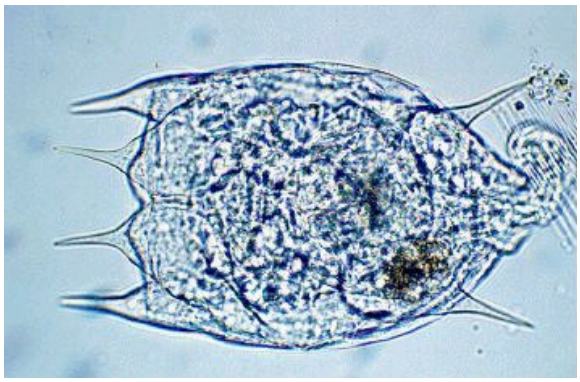

**Figure 4-6. *Brachionus calyciflorus* under the microscope.**

### Chydorid test

The Chydotox test (Pieters et al., 2008), using the benthic water flea *Chydorus sphaericus* (Figure 4-7), has been recently developed as a toxicity test for water extracts. Dekker et al. (2006) developed the culture conditions and the use of this organism for sediment testing. The endpoint of the Chydotox test is based on lethality.

The organisms are obtained from a laboratory culture. Juveniles at least 24 hours old are exposed to a range of concentrations of the water extract in 2 ml wide mouth crimp vials with seal in a volume of 250 µl. After 48 hours of exposure the dead or immobilized organisms in each test vessel are counted with the aid of an inverted microscope.

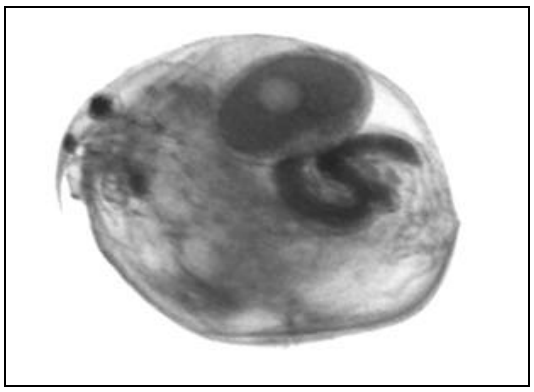

**Figure 4-7. *Chydorus sphaericus*, adult (photo: T. Dekker).**

The dose-response relation is estimated from the number of living, unaffected organisms in the concentrations in relation to the control. Toxicity values expressed as the concentration at which 50 % of the organisms are dead or immobile ($LC_{50}$/$EC_{50}$) are calculated using Toxcalc software (Tidepole scientific) or Graphpad Prism 4 (GraphPad Software Inc.).

## 4.3 Performance parameters for bioassays

This section gives information about the performance parameters for each bioassay. The selected bioassays in section 4.1.2 are derived from toxicity tests developed for estimating the toxicity of chemical substances. Most of the bioassays mentioned above are well standardized and have been validated in intra- and interlaboratory comparisons. Validation has not previously taken place using pure substances. A few studies are known on wastewaters and other environmental samples. Validation could be performed only on the basis of standardized methods. The test is considered standardized if detailed protocols are available. The test organisms should come from a common source and media and materials should be well-defined. After the bioassays had been standardized, it was possible to estimate their precision and sensitivity.

### 4.3.1 Precision of the bioassays

The precision of a method is a statement of the closeness of agreement between independent test results obtained under stipulated conditions (Prichard, 2006). Repeatability refers to the precision of measurements taken with the same method, in the same laboratory, using the same equipment and within a narrow time period. Reproducibility refers to the precision of measurements taken with the same method in different laboratories by different operators over a long time period. Precision is usually stated in terms of standard deviation(s); the relative standard deviation (RSD) or coefficient of variance; or the standard deviation of the mean (SDM) of a number of replicates. The reproducibility of a method often shows a larger variation than the repeatability, by as much as a factor of 2 or 3 (Prichard, 2006).

The precision of each bioassay is presented in Tables 4-3 and 4-4. The data for the bioassays have been collected from the literature or from experiments at RIZA or RIVM laboratories.

**Table 4-2. An overview of the repeatability of the bioassays listed above.**

| Bioassay | Repeatability (CV %) | Number of samples (n) | Sample | Reference (unpublished results marked with *) |
|---|---|---|---|---|
| Algae (PAM) | 2.8 | 40 | atrazine | RIVM results (2004)* |
| Algae (growth) | 1.2 | 3 | pure substances | Inter-laboratory comparison 2006 (results)* |
| Microtox® | 7-8 | 6 | effluent | Rotteveel and Roex, 2006 |
| Microtox® | 19 (2004)<br>19 (2005)<br>23 (2006)<br>15 | <br><br>6<br>16 | pure substances and effluents<br><br><br>phenol | Inter-laboratory comparisons 2004, 2005, 2006 (results)*<br><br><br>results (2005)* |
| Daphnia IQ | 3-32<br><br>6.5 – 10.1<br>34<br>9.7 | <br><br>3<br>20<br>20 | 11 pure substances<br><br>3 pure substances<br>potassium dichromate<br>potassium dichromate | Janssen and Persoone, 1993<br>Zaleska-Radziwill, 2000<br>RWS-RIZA results (2001)*<br>RWS-RIZA results (2005)* |
| Daphnia immobility | 20-27<br>16 | 6 | effluent<br>effluent | Rotteveel and Roex, 2006<br>Grothe and Kimerle, 1988 |
| Daphnia immobility | 14<br><br><br>5 – 14.4 | 3 | pure substances | Inter-laboratory comparison 2006 (RIZA-RWS results)*<br>ISO 6431 (1996) |
| Thamnotoxkit F™ | 0.4 – 1.1<br>13.5 | 3<br>10 | 3 pure substances<br>pure substance | Zaleska-Radziwill, 2000<br>Persoone, 1995 |
| Rotoxkit F™ | 16.8 | 7 | potassium dichromate | RWS-RIZA results (2005)* |

**Table 4-3. An overview of the reproducibility of the bioassays listed above.**

| Bioassay | Reproduci-bility (CV%) | Sample | Number of labs | Reference (unpublished results marked with *) |
|---|---|---|---|---|
| Algae (PAM) | - | - | - | - |
| Algae (growth) | 30 (2006)<br>37.7 | pure substances<br>pure substance | 5<br>16 | Inter-laboratory comparison (2006)*<br>ISO 8692 (2004) |
| Microtox® | 23 (2004)<br>27 (2005)<br>23 (2006)<br>18.2 | pure substances and effluents<br><br>pure substance | 11 (2004)<br>31 (2005)<br>29 (2006)<br>3 (n=81) | Inter-laboratory comparison (2004, 2005, 2006)*<br><br>Luoma and Ho (1993) |
| Daphnia IQ | 53 | copper sulphate | 16 | Aqua Survey information folder |
| Daphnia immobility | 54 (2006)<br>30 - 50<br>33 | pure substances<br>pure substances<br>effluent | 5<br>31 -36<br>9 | Inter-laboratory comparison (2006)*<br>ISO 6431 (1996)<br>Grothe and Kimerle (1988) |
| Thamnotoxkit F™ | - | - | - | - |
| Rotoxkit F™ | 48.5 | substances | 120 | Persoone et al. (1993) |

The repeatability expressed as the coefficient of variance (%) is for almost all toxicity tests less than 25 %. The reproducibility (also expressed as coefficient of variance) is higher and for most toxicity

tests in the range of 25-50 %. In comparison to other accepted methods these values seem to be acceptable.

### 4.3.2 Sensitivity/linearity/working range

The sensitivity of a bioassay refers to the change in response of a measurement divided by the corresponding change in stimulus. To give a measure of sensitivity, bioassay reference substances are routinely tested. The response of the reference toxicant is compared with the historical sensitivity data (Shewart control chart). Test results are rejected if the sensitivity of the reference toxicant differs significantly from the historical data. Table 4-5 shows the reference toxicants used for the different bioassays.

**Table 4-5. Critical boundaries of reference substances for the bioassays. Information on the number of observations (n) used to calculate the critical boundaries is limited and mentioned between brackets if available.**

| Bioassay | Reference substance | Critical boundaries ($EC_{50}$) | Reference (unpublished results marked with *) |
|---|---|---|---|
| Microtox® | phenol<br>phenol<br>potassium dichromate | 13 - 26 mg/l<br>14.3 - 25 mg/l (n=16)<br>18.7 mg/l | Microbics (1992)<br>RWS-RIZA (2005)*<br>ISO 11348-3 (1998) |
| Algae (PAM) | atrazine<br>atrazine | 55 - 80 µg/l<br>37 - 41 µg/l | Van Beusekom et al. (1999)<br>RIVM results (2004)* |
| Algae (growth) | potassium dichromate | 0.20 - 0.75 mg/l | ISO 8692 (2004) |
| Daphnia IQ | potassium dichromate<br>potassium dichromate | 10.1 - 38.6 mg/l<br>24.0 - 31.8 mg/l | RWS-RIZA (2001)*<br>RWS-RIZA (2005)* |
| Daphnia immobility | potassium dichromate<br>potassium dichromate | 1.12 mg/l | ISO 6431 (1996) (24 h) |
| Thamnotoxkit F™ | potassium dichromate<br>potassium dichromate | 0.067 - 0.088 mg/l*<br>0.044 - 0.118 mg/l | Microbiotest SOP (2006)*<br>RWS-RIZA (2002)* |
| Rotoxkit F™ | potassium dichromate<br>potassium dichromate<br>potassium dichromate | 9. 6 - 17.8 mg/l*<br>3.2 - 6.1 mg/l (n=5)<br>3.9 - 6.4 mg/l (n=7) | Microbiotest SOP (2006)<br>RWS-RIZA (2003)*<br>RWS-RIZA (2005)* |
| Chydorid | copper<br>potassium dichromate | 191 (184 - 198) µg/l (96 h)<br>285 (245 - 331) µg/l (48 h) | Dekker et al. (2006)<br>RWS-RIZA* |

* Specified for each batch culture

### 4.3.3 Additional remarks

During 2004 till 2006 the laboratory of RIZA has successfully participated in several internationally recognized laboratory comparisons for the toxicity tests with Microtox®, Daphnia and algae. The successfulness of a participant in testing a sample in an inter laboratory trial is expressed as a z-score. The z-score is a measure for the deviation from the mean produced by the participant in relation to the mean produced by all participants. A z-score is acceptable below 2. The z-scores for the toxicity tests performed at the RIZA laboratory were for all organisms < 2 (Ribo and Riva, 2004; Riva and Ribo, 2006).

# 5 Interpretation of results

## 5.1 Introduction

When the results of the toxicity tests are known, the question arises whether the results indicate if the location from which the tested sample was taken, is 'at risk' due to toxicity of substances. The basic assumption for interpretation of the results is that the original water sample should not cause chronic effects. The results from toxicity testing should be interpreted supplementary to Environmental Risk Limits. In contrast to Environmental Risk Limits, the presented methods for interpretation of the results do not intend to ensure an overall protection of the environment. Levels presented only give an indication whether chronic effects due to toxic substances, as far as they are concentrated, can be expected.

The acute toxicity caused by a concentrated sample may provide an indication of the presence of acute or chronic effects in the original sample. The applied Acute-to-Chronic ratio (ACR) is 10, which is based on a comparison of hundreds of tests for different species and substances (De Zwart and Sterkenburg, 2002; Raimondo et al., 2007). Chronic effects are therefore expected at concentration levels 10 times lower than the concentration where acute toxicity is observed. In section 5.4.7 the ACR ratio is evaluated.

Toxicity found in bioassays only represents the toxicity of substances that are actually concentrated from the original water sample. It depends on the properties of the substances to what extend they have been concentrated (see chapter 3). Therefore, it has been suggested in section 3.4.7 to use a safety factor of 0.5 for reduced recovery and substances which are hardly concentrated. With this extra safety factor the chance to neglect false-negative values is minimised.

Two methods are presented for interpretation of the results:

- judgement of results for single species and
- using risk analysis based on a species sensitivity distribution.

These methods are presented in the sections 5.2 and 5.3, respectively. In section 7.4 these methods are evaluated for the use of interpreting the results from toxicity testing.

## 5.2 Judgement of single species

Different test organisms from different functional groups can be used for toxicity testing on concentrates (chapter 4). The indication levels proposed in this section are intended for the evaluation of bioassay results using bacteria, algae and Daphnia, covering three functional groups in an aquatic ecosystem. Of course, other sets of tests can be used for this method. To compare different samples, the same toxicity tests should be used.

The results of the acute bioassays after concentration of the original sample are expressed as $EC^f_{50}$-values. These $EC^f_{50}$-values show the concentration factor of the original sample that causes an effect in 50 % of the test organisms ($EC^f_{50}$). Based on an ACR of 10, $EC^f_{50}$ value of 10 is thought to be the highest concentration factor at which chronic toxicity in the original sample is achieved. An $EC^f_{50}$ =10 means that in the original sample species are exposed at NOEC level.

In the chemical risk assessment an additional safety factor of 100 is applied to indicate the proportion for what is maximally permitted compared to what is negligible. This factor compensates e.g. for the potential effect of chemicals in combination (Traas, 2001). However, in the method presented in this report the organisms are already exposed to a combination of organic chemicals (as far as they are concentrated). Therefore a safety factor of 10 is used between the negligible and chronic level, instead of the factor of 100 applied in setting Environmental Risk Limits.

Two indication levels are proposed:

- indication of Chronic Effect (iCE), in at least one assay $EC^f_{50} < 10$;
- indication of Negligible Effect (iNE), within 3 assays $EC^f_{50} > 100$.

The assumptions used to establish these indication levels are:

When toxicity is encountered at levels below 10 times concentration, the iCE level has been achieved, since the results indicate that chronic toxicity is likely to occur.

These indication levels can also be expressed in Toxic Units, calculated as:

$$Toxic\ Unit\ (TU) = \frac{1}{ECf_{50}} \qquad \textbf{Equation 1}$$

These TU values indicate the dilution factor for the sample that causes 50 % acute effect in the assays used. Where the TU value is below 1, the sample needs to be concentrated to cause 50 percent effect. The indication levels for Chronic Effect and Negligible Effect are TU > 0.1 ($EC^f_{50} < 10$) and TU < 0.01 ($EC^f_{50} > 100$), respectively.

So far, these indication levels have not been corrected for the limited recovery of substances in the concentrated sample. To compare with these indication levels and take recovery into account as well, the toxicity results can be corrected for the limited recovery, e.g. by using the proposed safety factor from section 3.4.7 and 5.1. Taking this safety factor into account, the effect levels are $EC^f_{50} < 20$ (TU > 0.05) and $EC^f_{50} > 200$ (TU < 0.005) for iCE and iNE, respectively.

These proposed indication levels differ from the effect-based risk limits presented in Maas et al. (2003). This report presents effect-based risk limits to protect the aquatic environment from toxicity (chronic or otherwise) in the field. The reason that other values have been chosen is that in Maas et al. (2003) parameters have been used which are not reported in practice, like NOEC values for acute toxicity tests. Also, no indication level for serious effect is presented because it is not yet necessary to define such a level for water management practice.

## 5.3 Application of statistical extrapolation techniques

### 5.3.1 Introduction

Different species respond differently to a substance at a given concentration or even a water sample with a complex mixture of substances. These sensitivities can be reflected in a distribution, which is the basic assumption of species sensitivity distribution (SSD). According to Posthuma et al. (2002), an SSD is defined as a statistical distribution describing the variation among a set of species in response to

the toxicity of a certain substance or mixture. Available ecotoxicological data are seen as a sample of this distribution and can be used to estimate the parameters of the SSD. The SSD is estimated from a sample of toxicity data and visualized as a cumulative distribution function (CDF), which is the integral of an associated probability density function. This is shown in Figure 5-1. The cumulative distribution follows the distribution of species sensitivity data obtained from ecotoxicological testing.

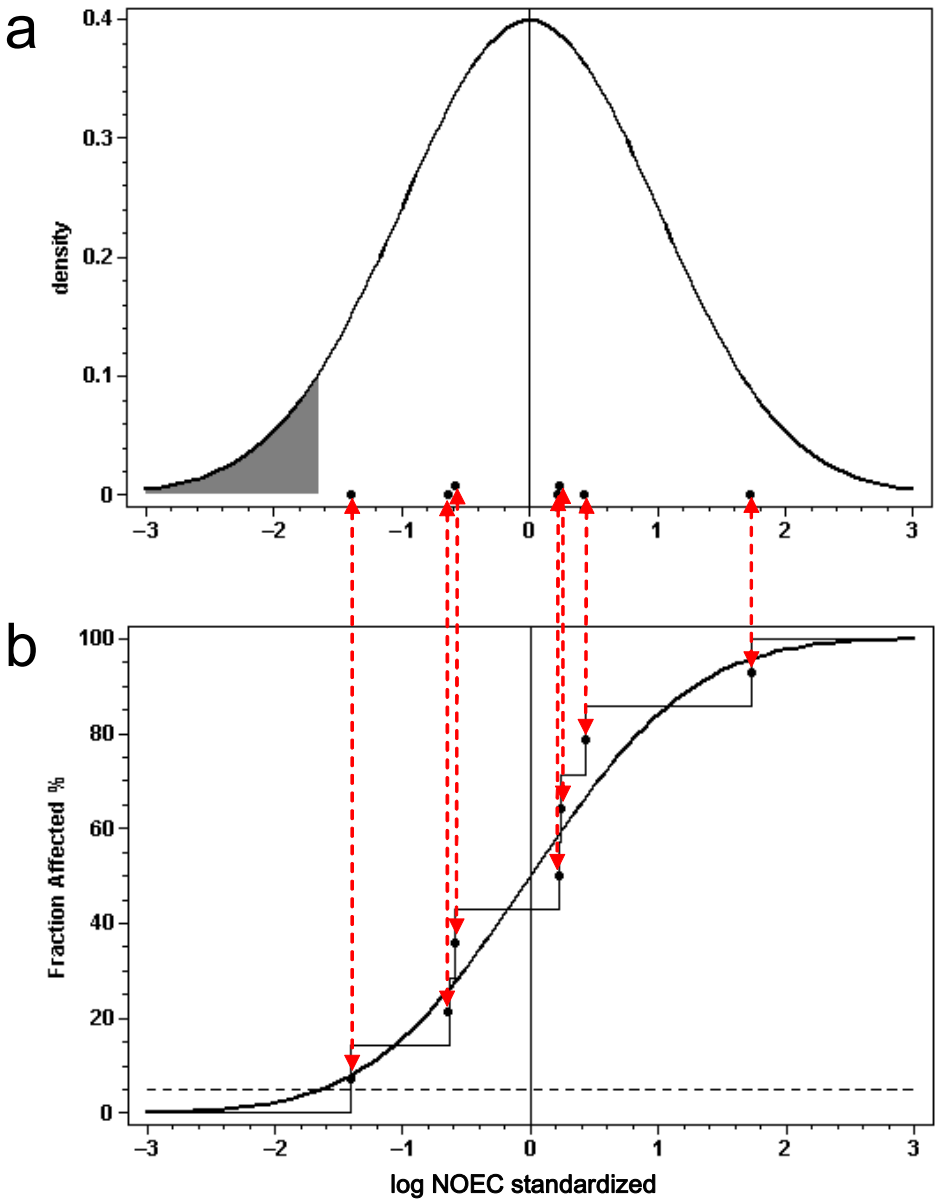


**Figure 5-1. Sample of toxicity data (dots) with the fitted probability density function (a) and cumulative density function (b). Red dashed lines denote the same toxicity data. The shaded area in curve a is the probability of selecting a species below the maximum permissible concentration, which is defined as the 5th percentile, shown as the grey dashed line in curve b.**

SSDs can be used in different ways. Originally these distributions were proposed to derive Environmental Risk Limits, like the maximum permissible concentration for the Dutch situation. For

this purpose, a concentration according a certain level of protection is derived from the distribution (i.e. the hazardous concentration for 5 or 50 percent of the species, called $HC_5$ or $HC_{50}$). This is known as the 'inverse use' of an SSD (Van Straalen and Denneman, 1989). 'Forward use' of an SSD is useful for ecological risk assessment, as it allows the fraction of the species in the community that is at risk to be estimated (Fraction Affected, or FA). Estimating FA is shown in Figure 5-2.

Different input data can be used to derive an SSD. The method presented in this report uses the results from a set of toxicity measurements. These measurements give an idea of toxic pressure of a certain water sample for the test species in the test battery. Using these measurements to derive an SSD, the fraction affected can be calculated indicating the toxic pressure for the test battery as a whole. This application is presented in section 5.3.2.

Metals are not concentrated by the XAD-technique, so the measured toxicity does not express the toxicity from different metals in the original sample. Fortunately, chemical analysis of metals can be performed quite effectively using a multi-element analysis technique like ICP-MS. This information can be used to gain an impression on the toxicity of metals, using knowledge of chemical speciation of metals and interactions of metals with biota (see Appendix III for proposed methods).

### 5.3.2 Calculating toxic pressure

De Zwart and Sterkenburg (2002) developed a calculating method for toxic pressure from measured toxicity in unknown mixtures. In this method, the endpoints of the toxicity measurements are used as a basis for calculating the toxic pressure of a water sample. The measurements are used to estimate the parameters of the SSD. The toxic pressure ('pT') of the untreated water sample is estimated from the SSD as the percentage of species exposed above their chronic NOEC. The observed acutely effective concentration factors ($EC^f_{50}$ or $LC^f_{50}$) are extrapolated to represent chronic no-effect concentration factors ($NEC^f$) by assuming a constant acute-chronic ratio (ACR) of 10 (see also section 5.1):

$$NEC^f = \frac{EC^f{}_{50}}{10} \; and \; NEC^f = \frac{LC^f{}_{50}}{10} \qquad \textbf{Equation 2}$$

Afterwards, the $NEC^f$ values are log transformed and fitted according to the chosen distribution function. Originally, De Zwart and Sterkenburg (2002) used the integrated log-logistic distribution function to fit the data. However, the log-logistic distribution can hardly be distinguished from the log-normal distribution for practical purposes. Because of the availability of standard functions in Microsoft Excel and the possibility to calculate confidence limits it is easier to use the log-normal distribution. Differences between these distributions will be discussed in section 5.4.4.

When we consider the toxicity measurements as a sample from a normal distribution, the mean and sample standard deviation are reasonable estimates of the parameters from a normal SSD. The cumulative distribution function can be estimated using mean and standard deviation. The fraction affected can be calculated using the Excel function NORMDIST(x, mean, standard_dev, 1), that returns the normal cumulative distribution for the specified mean and standard deviation:

$$pT = Fraction\ Affected = NORMDIST(^{10}LOG(C^f); AVG; STDEV; 1) \qquad \textbf{Equation 3}$$

Extrapolated no-effect concentration values (NECf) were $^{10}$log transformed before calculating the average log toxicity (AVG) and the associated standard deviation (STDEV). $C^f$ is the environmental concentration factor considered, where $C^f = 1$ corresponds to the original untreated water sample. This calculation is presented in Figure 5-2 graphically.

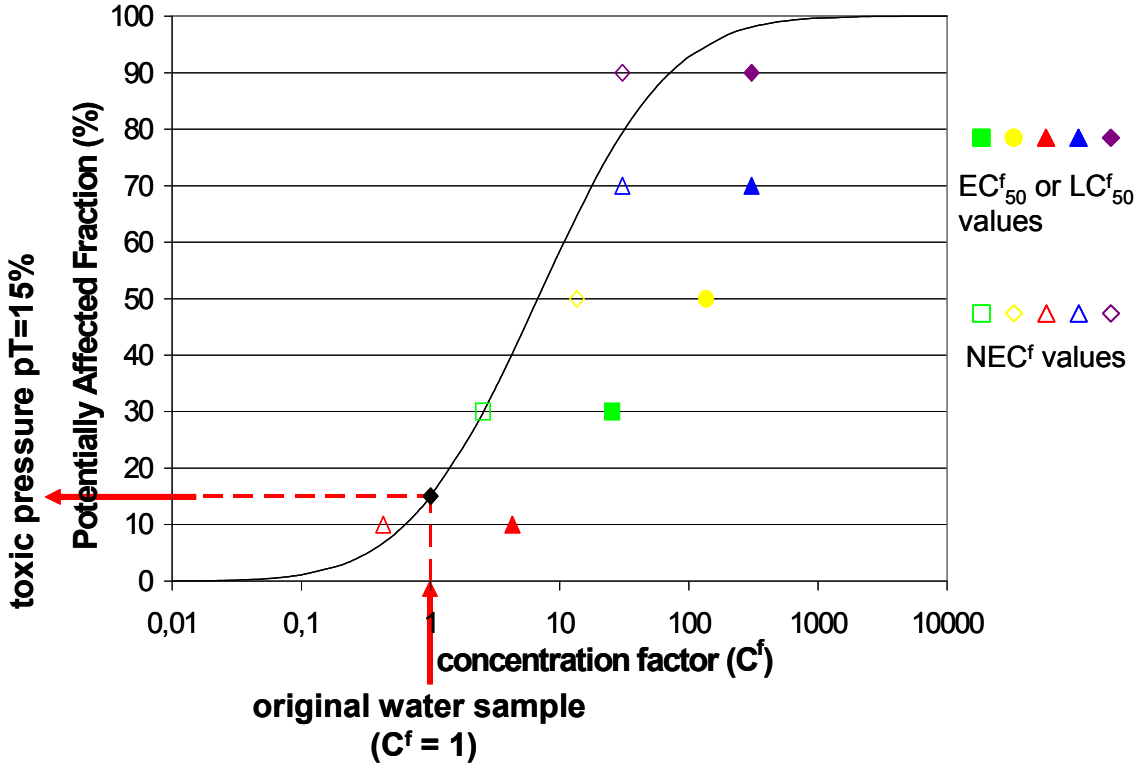


**Figure 5-2. Graphical presentation of the calculation of the toxic pressure of a water sample.**

An alternative for this single-fit distribution is the Bayesian approach in which the model itself is considered uncertain. Aldenberg and Jaworska (2000) analyzed the uncertainty of the normal SSD-model assuming unbiased species selection, and showed that both theories lead to numerically identical confidence limits. A salient feature of this extrapolation method is that the probability distribution functions and corresponding cumulative distribution functions are not determined as single curves, but as distributed curves (Aldenberg et al., 2002). Calculating Hazardous Concentrations or Fraction Affected as described in Aldenberg and Jaworska (2000) is implemented in the program $E_TX$ (Van Vlaardingen et al., 2004).

De Zwart and Sterkenburg (2002) developed the method for calculating toxic pressure with the results from 5 toxicity tests. In deriving Environmental Risk Limits in the Netherlands, a minimum of 4 toxicity tests (from at least 4 taxonomic groups) was considered required (Traas, 2001). To compare the toxic pressure for different locations, a minimum of 4 to 5 toxicity tests seems to be enough (chapter 7), assuming that the same tests are performed for the different locations. The influence of number of tests on uncertainty is evaluated in section 6.3.1 and 8.2.3.

The calculated pT value is very useful to compare toxic pressure of different locations. To judge the toxic pressure of a sample, two indication levels are proposed:

- indication of Chronic Effect (iCE), pT >5 % for $C^f = 1$;
- indication of Negligible Effect (iNE), pT <5 % for $C^f = 10$.

These proposed indication levels differ from the effect-based risk limits presented in Maas et al. (2003). Compared to this report another limit was chosen to indicate Negligible Effect, to keep the relation with the Species Sensitivity Distribution analogously to the indication of Chronic Effect Level instead of dividing iCE by a factor 100.

## 5.4 Evaluation

### 5.4.1 General

Toxicity testing is seen as an additional tool for monitoring water quality, effluent toxicity diagnosis, or finding the origin of toxicity. The test results give an idea of toxicity caused by known or unknown substances in the water. With the knowledge about the concentration procedure (chapter 3), it is obvious that not all potentially toxic substances are concentrated. This method gives information on toxicity, but only for the part of the substances in the sample which can be concentrated by XAD-4/8 and cause effects in the selection of tests performed. These limitations of the method should be taken into account for interpretation of results.

Monitoring results show that some test organisms are more sensitive to the toxic pressure of substances in water than others. However, it is not always the same bioassay that is most sensitive in every situation (see Table 5-1 or Maas and Van den Heuvel-Greve, 2005). From these results it can be concluded that the sensitivity of bioassays differs for toxic substances in water. Some tests turn out to be the most sensitive more often. It can not be predicted which species will be most sensitive to an unknown mixture of toxic substances. Therefore, different bioassays are needed to be able to show the toxic effects for the range of possible toxic substances in water.

**Table 4-1. Ranking of monitoring results since 2000 (1 = least toxic result, 5 = most toxic result), which gives an overview of the relative sensitivity of the different bioassays for the different samples (322 samples where at least these 5 bioassays were done)**

| Ranking no | Microtox® | Algae | Rotoxkit F™ | Thamnotoxkit F™ | Daphnia IQ |
|---|---|---|---|---|---|
| 1 | 1 | 5 | 73 | 225 | 43 |
| 2 | 2 | 15 | 174 | 71 | 37 |
| 3 | 76 | 36 | 60 | 19 | 129 |
| 4 | 164 | 85 | 9 | 5 | 59 |
| 5 | 79 | 181 | 6 | 2 | 54 |

### 5.4.2 Acute to chronic ratio

To test the criterion if chronic effects can occur in the original sample, the results from the toxicity tests are extrapolated to chronic toxicity using an acute-chronic ratio of 10. This ratio is used as well in deriving environmental risk limits (Traas, 2001; EC, 2003). The difference between acute and chronic toxicity varies widely from one substance or species to another (De Zwart and Sterkenburg, 2002, Raimondo et al., 2007). Since it is not possible to get an idea of chronic toxicity with short-term toxicity testing, acute toxicity measurements must be extrapolated to a value for chronic toxicity.

The acute-chronic ratio (ACR) of 10 used in the calculation of the toxic pressure of the water sample is arbitrary and rather conservative. The choice is based on the analysis of data from a study on the ecotoxicological quality of effluents from oil refineries, chemical industries and sewage treatment plants, where most of these ACR data were found to be below 10 (Figure 5-3).

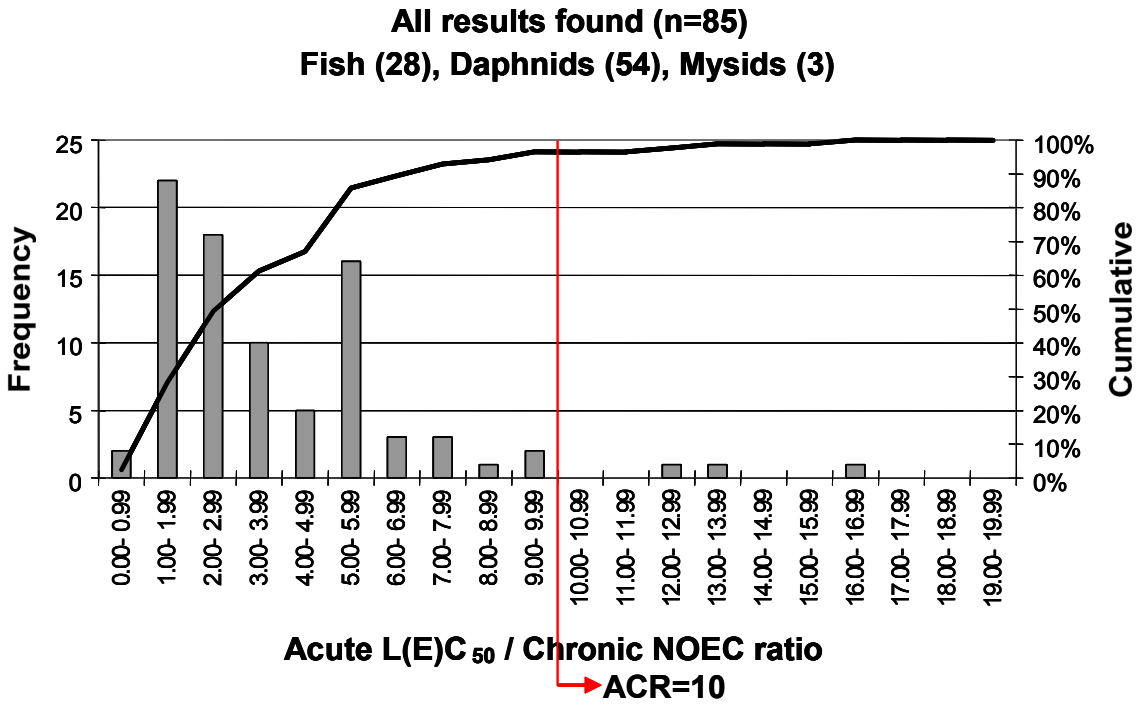


**Figure 5-3. Comparison of acute-chronic toxicity ratios for complex mixtures (De Zwart and Sterkenburg, 2002).**

For specific substances, it might be possible to calculate more specific acute-chronic ratios. But the method described in this report deals with possible complex mixtures of substances, and the ACR based on this data set can be considered as a 'best guess'. Considering the data, this would be a fairly safe guess in about 90 % of cases. In the other 10 %, the chronic toxicity will be underestimated.

More recent data analysis by Raimondo et al. (2007) results in a median ACR of 8.3. The 90$^{th}$ percentile of 79.5 for these data was obviously higher than for the data presented in Figure 5-3. Median ACRs for taxa, ambient habitat media, chronic test endpoint and chemical mode of action (MOA)/ class categories generally were similar but, in some cases, extremely variable. It should be taken into account that Raimondo et al. (2007) used data on separate toxic substances instead of mixtures.

### 5.4.3 Judgement of results

Chemical measurements can be judged using environmental risk limits. Risk limits should protect the ecosystem from adverse effects from the substance the risk limit is derived from (Traas, 2001). Therefore, data used for deriving environmental risk limits have to meet certain quality criteria and assessment factors are used dependent on the type and number of available toxicity tests for a substance. These assessment factors should provide sufficient protection for the ecosystem. The ecological meaning of toxicity measured in concentrated environmental samples is not clear, except that the toxicity tests that can be performed represent some important groups in the ecosystem. Chronic

effects on these important groups are undesirable. Statistical extrapolation of toxicity data is supposed to give a better foundation for judgement than judgement based on single species results, provided that enough toxicity data are available.

Because the ecological meaning of toxicity measured in concentrated environmental samples is not straightforward, it has been chosen to present indication levels instead of effect limits as presented in Maas et al. (2003). For the monitoring results since 2000, the judgement results were evaluated (Table 5-2). A judgement based on results for Daphnia, Microtox® and Algae (usually the most sensitive single species) exceed the indication of Chronic Effect level (iCE) less often than the judgement based on an SSD curve. The indication of Negligible Effect (iNE) is reached less often with the judgement based on the SSD curve than based on a set of three single species (usually the most sensitive).

**Table 5-2. Judgement of single species compared to results for judgement based on statistical extrapolation on monitoring data since 2000. Results are given as a number of cases above or below the mentioned indication level (344 cases in total). A safety factor of 0.5 is used for reduced recovery and substances which are hardly concentrated.**

| Judgement based on | >iCE | <iCE; >iNE | <iNE |
|---|---|---|---|
| Single species (Daphnia, Microtox® and Algae) | 43 | 267 | 34 |
| Statistical extrapolation | 64 | 267 | 13 |

Some examples for judgement on indicative effect levels are presented in Figure 5-4. For the toxicity data set where the iCE is exceeded obviously based on the SSD-curve (set A), the iCE level for judgement of single species is exceeded as well for two out of three single species taken into account for this judgement (Algae, Daphnia and Microtox®). For the cases where the iCE or iNE level is just exceeded when the judgement is based on the SSD-curve (toxicity data set B and D, respectively), these levels would not have been exceeded based on a single species approach.

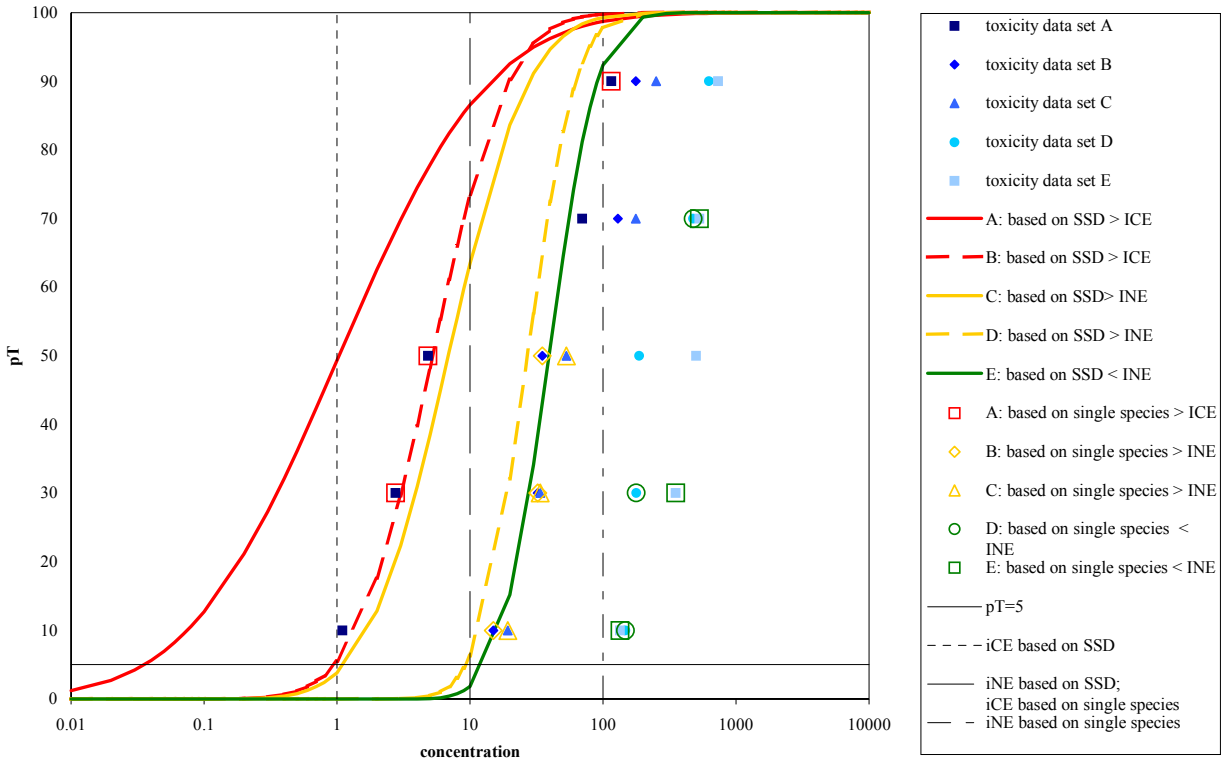


**Figure 5-4. Examples for judgement based on SSD-curves compared to judgement based on single species results for 5 cases.**

An important disadvantage of these judgement methods is that these chosen limits are always rigid. There is no difference between a small or a large exceedance, both situations give the same judgement. It might be more elegant to present these kind of results as a possibility or risk of exceedance of a certain level using Bayesian statistics.

### 5.4.4 Use of statistical techniques

Posthuma et al. (2002) mention that the SSD method requires three steps:

- selection of toxicity data;
- statistical analysis of those data;
- interpretation of the output.

The output is of course very dependent on the choices made in each of these steps. An evaluation of the consequences of the choices made is presented in this section.

**Selection of toxicity data**

In Traas (2001) a set of four taxonomic groups was considered enough to apply statistical extrapolation methods (SSD). In the European Technical Guidance Document (TGD) (EC, 2003), the criteria to apply this method for deriving a predicted no effect concentration (PNEC) are much more rigid, with at least 10 NOECs (preferably 15) covering at least 8 taxonomic groups. Dependent on the data set, an additional assessment factor between 5 and 10 should be applied to derive a PNEC. Because the test set for toxicity testing on concentrated environmental samples can never meet the TGD criteria (used for

setting Environmental Risk Limits), interpretation using SSDs might give a less safe judgement of ecosystem effects compared to TGD criteria.

**Statistical analysis of the data**

There is no good basis for selecting a distribution function to fit the data, especially when the number of observations is small (Posthuma et al., 2002). Like many other uses of SSD, the chosen distribution for this method is the same as the distribution chosen for another purpose: in this case the derivation of quality objectives in the Netherlands. The logistic probability function has been designed to describe resource-limited population growth (or in this case: toxicity limiting population growth) and should by its nature be capable of describing the natural situation. Another argument for using a logistic probability distribution rather than a normal probability distribution is that the log-logistic distributions render marginally more conservative values in the estimation of hazard concentration (De Zwart, 2002). Compared to the log-normal distribution, the log-logistic distribution will also lead to more conservative FA values.

In literature, several comparisons between distribution functions are made, but where one finds the lognormal distribution consistently preferable for one dataset, the other finds the log-logistic distribution gives a better fit. In fact, several distributions give quite similar results. Other issues are likely to be more important, such as the choices made in data treatment, related to ecological issues. (Suter et al., 2002)

Goodness-of-fit calculation for the log-logistic or log-normal distribution gives comparable results. Using the log-normal distribution is more practical, because of the availability of a standard function in MS Excel.

The SSDs for bioassays tend to be much steeper than those for individual chemicals. De Zwart and Sterkenburg (2002) suggest a few hypotheses to explain this phenomenon. Since it is unlikely that the bioassay species are all extremely sensitive (or insensitive) to some of the chemicals in the mixture, the response to a complex mixture tends to be closer to the median than the response to individual chemicals. Some species are very sensitive to a specific single chemical, which strongly influences the slope of the SSD. Another hypothesis is that all single chemicals in the cocktail act as a single virtual toxicant with a narrow action spectrum.

**Interpretation of the output**

Although toxicity tests can give an idea of the combined effect of toxic substances, several reasons are mentioned why interpretation of toxicity tests on concentrated environmental samples might not give a safe judgement on ecosystem risk, compared to Environmental Risk Limits. Not all potentially toxic substances are concentrated and the recovery of concentrated substances strongly depends on substance properties. It can therefore never give a complete view on toxicity of a sample. Environmental Risk Limits are derived to give a safe judgement for every substance, but only for substances measured chemically and not for actual combined toxic effects. The possible test set for toxicity testing on concentrated environmental samples is also rather limited and can never meet the TGD criteria (10 NOECs, 8 taxonomic groups), which makes the results of the method presented in this report not suitable enough for a safe judgement on ecosystem risk.

Nevertheless, SSDs can be used very well for estimating the Fraction Affected if the same bioassays are used (Posthuma, 2002) and the same concentration method is applied. Comparison with a Fraction Affected calculated with another set of test species is not recommended. Dependent on test species considered, the SSD can change quite strongly. It can be questioned whether the available test set

describes the possible variation among the species in the field in response to the toxicity of the mixture of substances in the sample.

# 6 Validation of the entire method

## 6.1 Introduction

This chapter evaluates the different performance parameters for Quality Assessment and Quality Control (QA/QC, see section 6.2) that might be necessary for the entire method. For some of the performance parameters results are available. Other parameters are not properly worked out and are subject of discussion. Validation has to be carried out within a single laboratory (intra) and between different laboratories (inter). The validation results in this chapter are mostly obtained from single laboratory studies during a period of several years. Only one comparison could be made between different laboratories.

The key performance parameters that require attention during validation vary with the purpose for which the method is used. Table 6-1 shows different types of analysis and the performance criteria that require evaluation to demonstrate that a method is fit for purpose.

**Table 6-1. Performance parameters required for validation of different types of analysis (Prichard, 2006).**

| Performance parameter | Type of analysis for water data information | | | |
|---|---|---|---|---|
| | Qualitative | Major component | Trace analysis | Physical property |
| Precision | | + | + | + |
| Specificity/selectivity | + | + | + | + |
| Bias/recovery | | + | + | + |
| Ruggedness | + | + | + | + |
| Sensitivity/linearity/working range | | + | + | + |
| Limit of detection | + | | + | |
| Limit of quantitation | | | + | |

The XAD-method as described in this report is comparable to the trace analysis, but several of these performance parameters are difficult to estimate for the XAD-method in total. The performance criteria mentioned above are derived for analytical measurements. In the XAD method, however, the target for biological measurements is difficult to define.

For the implementation of the Water Framework Directive (WFD, 2000/60/EC) the EU Commission has made a guidance on surface water chemical monitoring under the Water Framework Directive (Concept-version 9, Lepom et al., 2008). This Guidance provides general rules for sampling, sample treatment and analytical methods to be used in the surveillance, operational and investigative monitoring of the WFD. In this Guidance minimum performance criteria for analytical methods for chemical monitoring are specified next to other parameters relevant to compliance checking. These criteria have been laid down in the Final Draft of technical specifications for chemical analysis and monitoring of water status (Quevauviller, 2008).

Minimum performance criteria have been defined as the Limit Of Quantification (LOQ) and the measurement Uncertainty (U). They are linked to the Environmental Quality Standards (EQS) where

possible. In case an analytical method may not meet these minimum criteria laid down in the Decision paper from the European Commission (Quevauviller, 2008), the best available technique should be carried out not entailing excessive costs. The Guidance describes the methods how to estimate the minimum criteria in a pragmatic way.

In section 6.2 an explanation is given for the performance criteria and the results of the quality parameters that could be applied on the XAD-method. The interpretation of the results is evaluated in section 6.3

## 6.2 Performance parameters

For the XAD-method including bioassays it is tried to join the QA/QC performance criteria mentioned in the Decision paper (Quevauviller, 2008). The XAD-method on environmental samples was difficult to compare with the method for analytical analysis. For analytical analysis it is clear what target has to be measured (for instance, to meet the EQS for a priority pollutant). For the method based on results of environmental samples it is hard to estimate those performance criteria. Bioassays do not specifically target individually substances. The technique gives a single holistic answer, the overall summation of the toxicity of all substances present in the sample. Prichard (2006) has specifically written a validation guideline that deals with the requirements, problems and the validation of emerging tools and possible ways of establishing performance criteria. She mentioned a possible validation approach by validating the method with a small number (4-5) of agreed relevant individual target substances. However it will take a lot of effort to measure these criteria with known samples. Besides, every analytical or environmental sample will react in a different way. This is a common phenomenon for validation studies on bioassays. So, for several performance criteria a pragmatic approach has been suggested. In this section the following criteria will be discussed:

- standardization of the method;
- precision of the entire method;
- specificity / selectivity;
- ruggedness;
- limit of detection / limit of quantification;
- measurement uncertainty.

### 6.2.1 Standardization of the method

Implementation of the method for monitoring, for instance under the WFD, requires a well-defined description of the entire method. On basis of the experiences with the method during several years of monitoring, defined protocols from sampling through interpretation of the results have been drawn up. Wherever possible the methods according available international standards are used. The detailed protocols are given in Appendix I.

### 6.2.2 Precision of the entire method

The precision of a method is a statement of the closeness of agreement between independent results obtained under stipulated conditions (Prichard, 2006). Two types of precision could be estimated: repeatability and reproducibility. Repeatability refers to the precision of measurements taken with the same method, in the same laboratory, with the same equipment and within a narrow time period. Although not many data are available, the studies suggest that the repeatability of the total procedure lies within acceptable ranges. An initial validation study with effluent samples (Rotteveel and Roex, 2006) indicated that the variation between toxicity measurements on a sample from one extraction run is only slightly lower than the variation between toxicity measurements on samples from different

extraction runs (Table 6-2). The variation in the entire method seems to be determined mostly by the variation of the toxicity methods.

**Table 6-2. Repeatability of the total procedure for effluent samples, given for the algae growth test, the bacteria test (Microtox®) and the acute daphnia test. The effects are measured on four (n=4) samples from different extraction runs of the same sample (three effluents studied). Some bioassays are repeated six (n=6) times on the same extract (two effluents studied).**

| Bioassay | Repeatability (CV %) | | | | | Reference |
|---|---|---|---|---|---|---|
| | total method repeated (n=4) | | | tests repeated in one concentrate (n=6) | | |
| | Efl 1 | Efl 2 | Efl 3 | Efl 1 | Efl 2 | |
| Algae (growth) | - | - | 19 | - | - | Rotteveel and Roex, 2006 |
| Microtox® | 11 | 11 | 3 | 8 | 7 | |
| Daphnia | 41 | 30 | - | 27 | 20 | |

**Repeatability**

A comparable study has been applied for samples of surface water (unpublished data RIVM and RIZA, 2002). Three samples (n=3) are taken at the same moment on two locations. The samples were treated independently and the toxicity is measured with a battery of five test organisms (Table 6-3). The variation was comparable with the effluent study (Table 6-2).

**Table 6-3. Repeatability of the total procedure for surface water, including sampling, given for the bacteria test (Microtox®), the Daphnia IQ test, the Thamnotoxkit F™ and Rotoxkit F™.**

| Bioassay | Repeatability (CV %) | | Sample | Reference |
|---|---|---|---|---|
| | total method repeated (n=3) | | | |
| | Loc 1 | Loc 2 | | |
| Microtox® | 3.3 | 13.3 | Surface water from agricultural area | unpublished data RIVM and RIZA, 2002 |
| Daphnia IQ | 18.1 | 10.4 | | |
| Thamotoxkit F™ | 22 | 7.8 | | |
| Rotoxkit F™ | 15.9 | 34.1 | | |
| Algae (PAM) | 14 | 7.3 | | |

**Reproducibility**

The reproducibility (precision of measurements taken with the same method in different laboratories by different operators) of the complete procedure has been applied for one location of surface water. Three laboratories have extracted and concentrated seven times the same field sample at the same time. The sample was sampled on an agricultural area and divided in 21 subsamples. Each laboratory treated seven sub samples following the XAD-procedure. Bioassays with daphnia, Microtox® and algae (plate test) were applied on the 21 extracts by one laboratory. The results of the entire procedure for two laboratories was almost the same, the third one differed from the other two laboratories (Table 6-4; Figure 6-1).

A second test on reproducibility has been performed on a technical mixture of different compounds. The three laboratories have concentrated the organic fraction and the recovery of the compounds added will be analysed. Results will be presented by Roex et al. (in prep.).

**Table6-4. Repeatability of the total procedure for surface water, inclusive sampling, given for the algae (plate) test, the bacteria test (Microtox® ) and the Daphnia test (3 labs; n=7)**

| Bioassay | Repeatability (CV %) (n=7) | | | Sample | Reference |
|---|---|---|---|---|---|
| | Lab 1 | Lab 2 | Lab 3 | | |
| Algae (growth) | 33 | 35 | 20 | Surface water from agricultural area | Roex et al. (in prep.) |
| Microtox® | 15 | no toxicity found | 18 | | |
| Daphnia | 5 | no toxicity found | 6 | | |

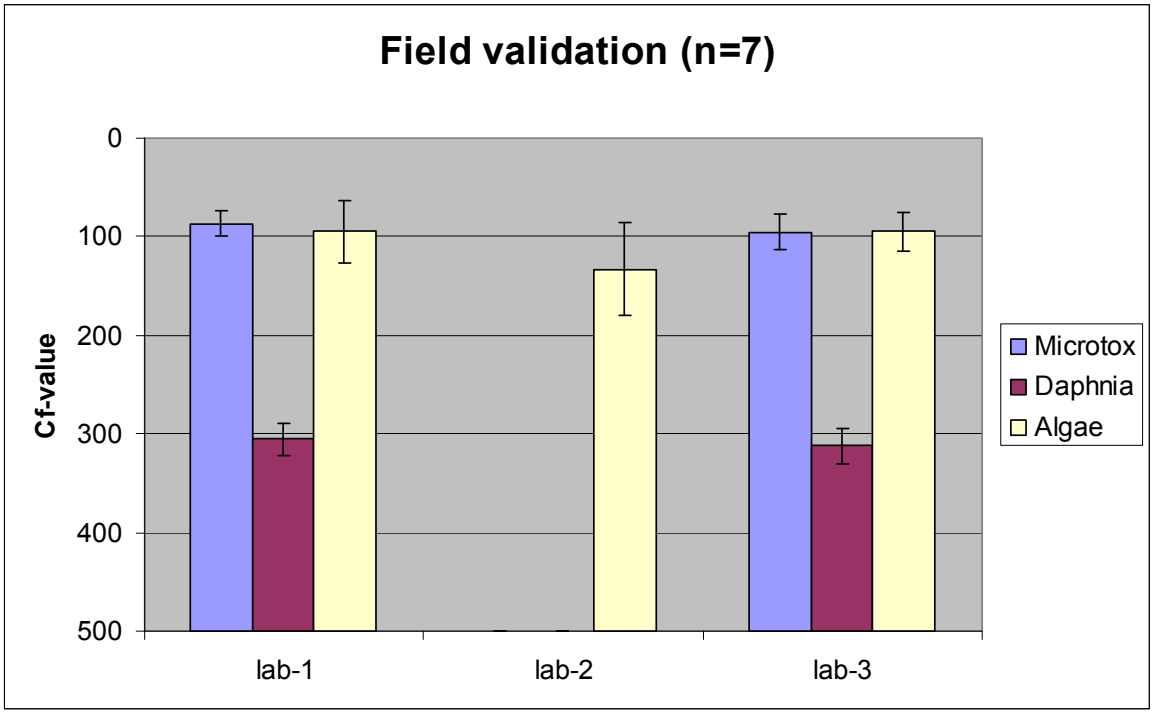


**Figure 6-1. Results of the inter laboratory test with a surface water sample. The samples have been extracted in 7 replicas at three laboratories. The toxicity tests are applied at one laboratory. Error bars show the standard deviation (SD).**

### 6.2.3 Specificity/selectivity

The specificity of a method refers to the extent to which it can detect and determine a particular analyte in a mixture, without interference from other substances in the mixture. This means that a bioassay should detect toxicity caused by pollutants and not by confounding factors like low oxygen levels, high pH or high salinity. Postma et al. (2002) have given ranges for confounding factors for most of the bioassays used for this method. However, these confounding factors are not of concern after extraction of the sample, because the extraction procedure is selective for organic pollutants. These pollutants are dissolved in a standard biological test medium afterwards. Metals and other inorganic substances are not concentrated. Extraction therefore also precludes toxicity as a result of confounding factors.

The extraction of organic pollutants may be affected by certain factors. In chapter 3 the influence of different factors has already been discussed. Standardization of the method should prevent variation in results due to experimental circumstances.

The influence of the factors suspended solids and humic acids on the entire method were investigated. XAD was added to untreated water samples (3.4.2). The presence of suspended solids during the XAD extraction did not show much influence on the concentration of toxicants in the extracts. Only small but no significant increases in toxicity were observed in the untreated fraction in comparison to the filtrated fractions (Figure 6-2).

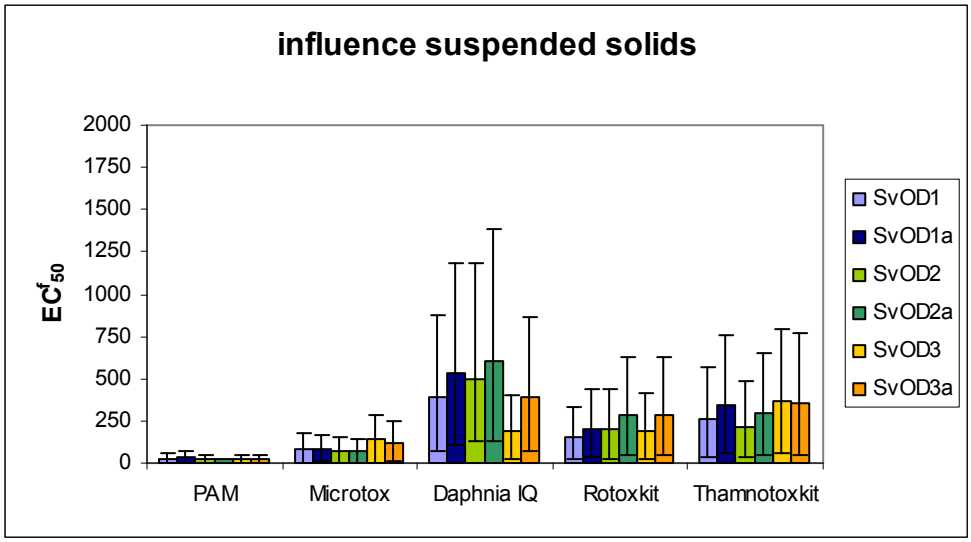


**Figure 6-2. Effect of filtration on the toxicity of samples from location 'Schaar van Ouden Doel (SvOD)' in the river Scheldt. The samples notated with 'a' are filtered and can be compared with samples with the same number. Error bars show the confidence intervals.**

The presence of *humic acids* in surface water did not affect the toxicity of several substances either. A comparison of the entire procedure for substances spiked to surface water (including humic acids) and spiked to standard test water showed the same toxicity (see section 3.4.2).

### 6.2.4 Ruggedness

The ruggedness of a method is evaluated by measuring the changes in response after small adjustments in the measurement conditions. For example, the pH, salinity and other factors might be altered and changes in the response of the assay determined. The aim is to identify and, if necessary, better control method conditions that might otherwise lead to variation in the measurement results, when measurements are carried out at different times or in other laboratories. It can also be used to improve the other performance parameters of the method.

The concentration procedure and most bioassays have been evaluated for ruggedness. However, the changes in the total response after changes in the original sample to these factors mentioned above have not been investigated. To prevent variation between laboratory results the protocol for the extraction and concentration procedure should mention the experimental conditions for these factors (temperature, light conditions). Although the pH and the salinity also will vary, these parameters are not changed

during the extraction procedure, in order to keep the chemical availability as relevant as possible for the estimated location.

Bioassays are carried out under restricted circumstances, laid down in the protocols or international guidelines. Sensitivity of the test organisms is always tested in comparison to a response of reference substances, also laid down in the protocols. As already said in section 6.2.3, confounding factors like pH, nitrate, salinity etc. are not of concern after the extraction procedure of the sample. Other factors like temperature and light conditions are laid down in the protocols of the bioassays.

### 6.2.5 Limit of detection (LOD) and limit of quantitation (LOQ)

The minimum level that can be detected is denoted by the Limit Of Detection (LOD), which is determined by measuring the variation in response of samples without pollution. The LOD is fixed at three times the standard deviation of the blank, corrected for the mean level of the blank itself. In general, the noise of an analytical measurement is taken as response level of the blank.

$$LOD = Mean_{blank} - \left(3 \times SD_{blank}\right) \qquad \textbf{Equation 4}$$

The Limit Of Quantitation (LOQ) is defined as the lowest concentration of analyte that can be determined with an acceptable level of uncertainty. It should be established using an appropriate measurement standard or sample. It should not be determined by extrapolation. Various conventions take the limit to be 5, 6 or 10 standard deviations from the blank measurement (Prichard, 2004). By Quevauviller (2008) a factor of 9 is used for minimum performance criteria for analytical methods,.

Mineral water for human consumption (Spa) is often used to measure the toxicity of the blank procedure. The variation in response in the blank extracts should be used to calculate the limit of detection and the limit of quantitation. The response of all organisms is measured in a blank procedure for several times. For most of the responses the $EC^f_{50}$-value was higher or in the neighbourhood of the maximum concentration tested (varying from 500 - 1000x). Sometimes the $EC^f_{50}$-values are extrapolated and are measured above the concentration factor of 1000x. As a consequence the standard deviation of the $EC^f_{50}$ is very high. To define proper LOD and LOQ values for the responses of the organisms the response value of the blanks should be determined more accurately.

Another way to deal with this problem could be a more pragmatic approach. In our opinion it is only necessary to estimate the concentration factor of the blank in which no toxicity is found in the different bioassays. The minimum response on blanks from results of five different bioassays over the last five years was at a concentration factor of 300 ($EC^f_{50} > 300$). For several bioassays this response is 400 or more.

What does this level mean?
Responses in the neighbourhood of the concentration factor of 300 will be less certain. However responses at $C^f$=300 are not relevant for the ecosystem. In chapter 5 the indication for Negligible Effect is set on 100, indicating that no effects are expected above that value. Taking a safety factor of 0.5 into account (for recovery and substances that will hardly be concentrated), responses with $C^f > 200$ are considered to indicate a negligible effect.

### 6.2.6 Measurement uncertainty

Measurement uncertainty is a parameter associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the quantity that is being measured

(VIM, 1996). In analytical chemistry it typically encompasses contributions from sample processing and treatment (e.g. extraction of the analyte) as well as from instrumental readings.

A target uncertainty means the measurement uncertainty that is formulated as a goal and that is decided on the basis of a specified intended use of measurement results (e.g. like EQS for priority pollutants). The minimum required relative target uncertainty for the priority pollutants is set on 50 %. Knowledge of the magnitude of the contributing errors from each step or process in an analytical method can be helpful to identify the significant errors to address if action is desirable to reduce the overall measurement uncertainty (Lepom et al., 2008). For the method presented here (sampling, sample treatment and bioassays), such an approach is not feasible because the composition of the samples by definition is not known.

An alternative approach of estimating the measurement uncertainty is to use data from analysis of certified reference material, routine control samples or inter laboratory trials. For the XAD-method a lot is already known about different mixtures of substances (chapter 3) and blanks. The inter laboratory trial with a field sample gives an indication of the variation in test results. Important for this criteria is the uncertainty at the level of the concentration factor of the critical value ($EC^f_{50}$ = 10). To measure the uncertainty at this level further intra and inter laboratory surveys should be performed with a known mixture of substances with a response at the critical level. However every mixture of known substances will react in a different way as a field sample. The response in the bioassays of the field samples analysed in the inter laboratory trial were approximately at the level of $C^f$ = 100. The measurement uncertainty (CV) at this concentration varied between 5 – 35 %.

## 6.3 Interpretation of the results

### 6.3.1 Ruggedness of pT

Aldenberg et al. (2000) tabulated upper, median and lower estimates for the Fraction Affected for standardized concentrations in relation to toxicity data sample size. The interval between the lower and upper estimate (5 % and 95 % confidence respectively) decreases slightly with increasing number of toxicity tests (sample size, n). The standardized concentration (-ks), calculated as:

$$-ks = \frac{log(C^f) - mean}{standard\ deviation} \qquad \textbf{Equation 5}$$

has a stronger relation with the confidence interval. This means that the difference between the original water sample ($C^f$ = 1) and the mean of the SSD-curve determine the confidence interval together with the standard deviation of the samples. Increasing the number of toxicity tests in the sample to estimate an SSD has therefore only little effect. Only very large sample sizes decrease the confidence interval substantially.

Standardized concentrations are calculated for monitoring data from 2000-2005 (Figure 6-3). Also for this data, based on calculations for a single fit distribution with Excel, the small effect of increasing the number of toxicity tests (or sample size = n) on the confidence interval is obvious.

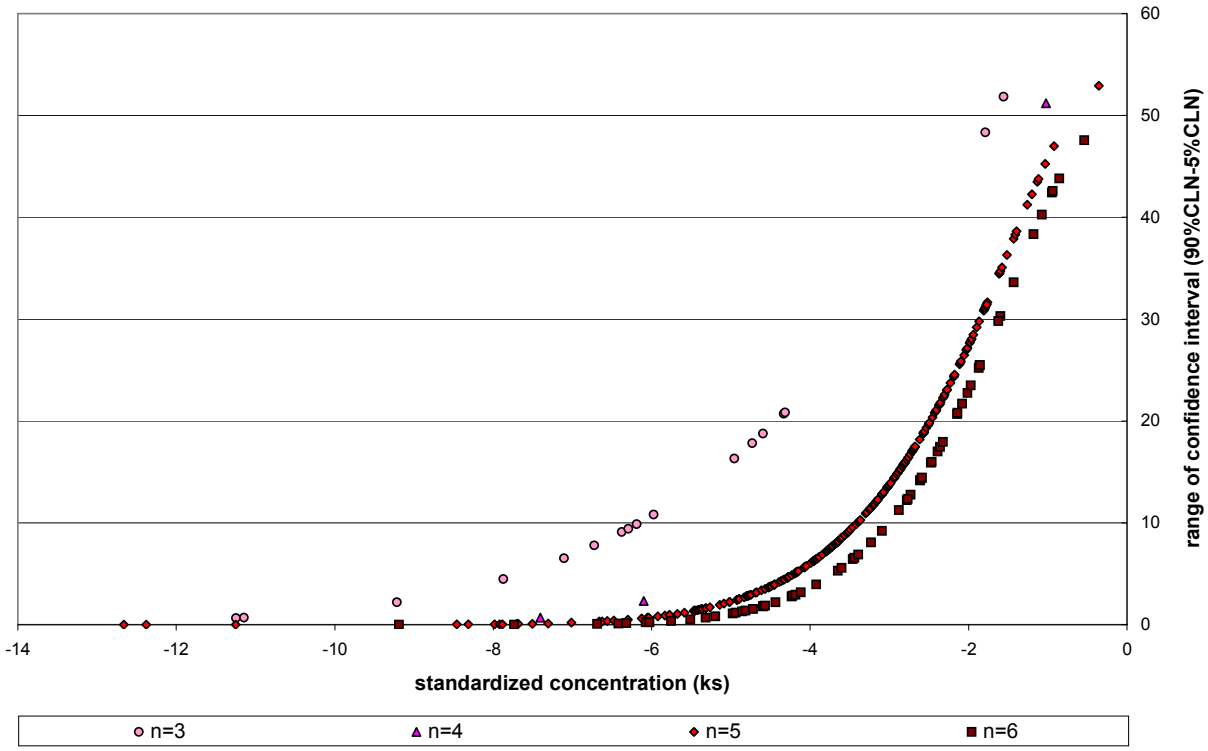


**Figure 6-3. Relation between confidence interval (range) and standardized concentration (ks) for different numbers of toxicity tests (sample size = n).**

### 6.3.2 Repeatability of pT

For three samples from agricultural area repeated measurements have been done for a test battery of five bioassays. The differences in coefficients of variation for toxic pressure (pT) are very large.

**Table 6-5. Repeatability of pT for three samples from agricultural areas.**

| Sample | pT | 5 % | 90 % | average | SD | CV (%) |
|---|---|---|---|---|---|---|
| ECO-5a | 8.2 | 0.8 | 39.4 | 8.5 | 2.003731 | 23.54572 |
| ECO-5b | 10.7 | 1.4 | 42.9 | | | |
| ECO-5c | 6.7 | 0.5 | 37.2 | | | |
| ECO-6a | 0.4 | 0.0 | 17.3 | 0.2 | 0.12799 | 58.42099 |
| ECO-6b | 0.2 | 0.0 | 15.0 | | | |
| ECO-6c | 0.1 | 0.0 | 12.9 | | | |
| ECO-27a | 0.5 | 0.0 | 18.9 | 0.6 | 0.070839 | 12.5757 |
| ECO-27b | 0.6 | 0.0 | 19.7 | | | |

The values for pT will vary with the parameters of the SSD, the average and the standard deviation of the test results. As Aldenberg et al. (2002) showed, the estimated Fraction Affected is more sensitive for variation in the most sensitive or most insensitive species, than for variation in species for which toxicity is close to the average of the distribution. In other words: if the variation in toxicity for a very sensitive species is very large, it will affect the value for toxic pressure (pT) more strongly than if the variation in toxicity in an average species is large.

## 6.4 Summary

Although a lot of Quality Assessment and Quality Control has been performed for the different parts of the method, the validation of the entire method is not fully completed yet. In chapter 8 recommendations to improve the most important performance criteria will be given.

# 7 Applications of the method

## 7.1 Whole Effluent Assessment

It is generally recognised that the substance-by-substance approach has some shortcomings. Results from chemical analyses of wastewater samples have shown that only a limited number of substances can be analysed, identified and/or quantified. Furthermore, environmental data (persistency P, bioaccumulation B, and toxicity T) are often lacking for a substantial part of the substances eventually identified in an effluent.

This is one of the reasons for the ongoing interest in the development and implementation of biological tests that can be applied to entire environmental samples like effluents. These tests have already shown that the substances identified can only partly explain the measured adverse effects. This means that a large fraction of the adverse biological effects in effluents is caused by 'unknown' substances or by the mixture of substances.

Whole Effluent Assessment (WEA) can be defined as the assessment of effluents using a range of biological methods (P, T, estrogenic activity) and chemical analyses (B) in order to reveal potential PBT effects. The outcome of the WEA tests can be used to select/determine and assess the Best Available Technique (BAT) on an individual basis. In the Netherlands, WEA parameters have not so far been used in licensing (for discharge), though it is expected that the use of WEA parameters for setting Emission Limit Values (ELVs) as well as for monitoring discharges and enforcing permits will increase, especially in the context of IPPC (Integrated Pollution Prevention and Control).

WEA parameters are not yet used to assess the impact of effluents on the receiving water. However, besides its application in licensing, WEA might play a useful role in the implementation of the Water Framework Directive (WFD). It is expected that, in the near future, WEA will be used for management planning under the WFD. Where a water body has poor ecological status and the suspected cause is discharges of unknown chemical substances, WEA can be used to identify and prioritise relevant point sources. WEA tests for acute toxicity, bioaccumulation and endocrine disruption will be carried out in surface water and effluents.

The Dutch WEA toolkit includes tests for acute toxicity (bacteria, algae, crustacean) and endocrine disruption (ER Calux bioassay). Today's standard procedure for acute toxicity testing in the Netherlands is toxicity testing after XAD extraction/concentration (Roex and Rotteveel, in prep.). The advantage of this procedure is that a near effect value can be generated. This is extremely helpful when assessing the effectiveness of a new or existing treatment to estimate the impact of an effluent on the receiving surface water.

## 7.2 Monitoring

From 1996 to present a monitoring campaign is carried out by RIVM and the Centre of Water Management (formerly RIZA) together within the program 'Development and application of bioassays to judge water quality'. Surface water is sampled regularly to measure toxicity on several locations in the Netherlands. The rivers Rhine and Meuse are sampled yearly every two months on the locations

Lobith and Eijsden, where these rivers cross the borders with Germany and Belgium respectively. Some years, other locations were sampled as well every two months. Within this monitoring program, water samples were concentrated. With the concentrated water samples, five different toxicity measurements were carried out using the Microtox®assay, Algae PAM test, Daphnia IQ test, Thamnotoxkit F™ and the Rotoxkit F™.

After 2005, the results from the past 10 years of monitoring water quality with bioassays were evaluated (De Groot et al., 2007. The monitoring results showed an obvious break in trends between 1999 and 2000 (Figure 7-1), which is caused by an improvement in the concentration procedure resulting in an improved recovery for narcotic substances (Struijs and Van de Kamp, 2001) and reduced introduction of toxicity (Vaal et al., 1999). This report only presents the improved method (see chapter 3). Details on the concentration procedure used before 2000 can be found in Collombon et al. (1997).

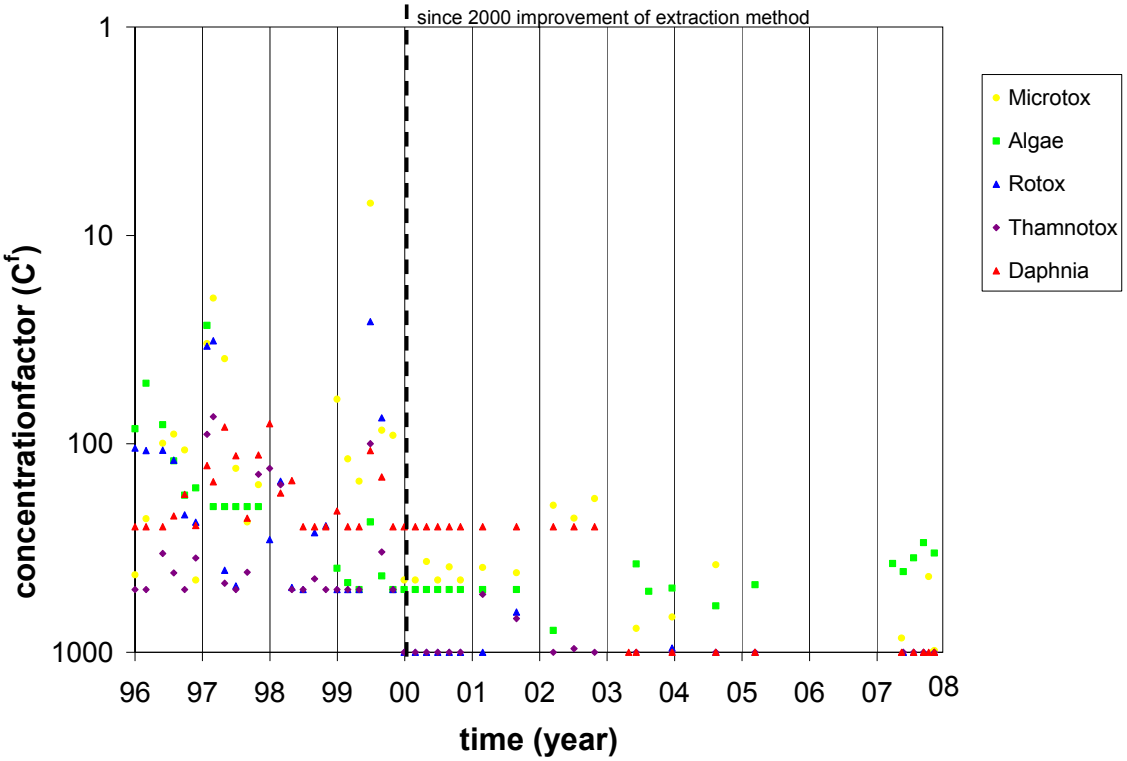


**Figure 7-1. Results of toxicity tests on blank samples (mineral water) from 1996-2007. For Daphnia, the highest concentration factor tested has changed since 2003 (Cf=250 before 2003; Cf=1000 after 2003)**

The toxic pressure (pT) was calculated for all monitoring data since 2000, and resulted for most samples in an average pT value of 0 %. This means it can be assumed with high certainty that concentrated toxic substances don't cause toxic pressure. The highest pT values were calculated for Eijsden, followed by location Schaar van Ouden Doel in the river Scheldt which was sampled less regularly. The average pT-values in the Meuse on location Eijsden for the period 2000-2007 is shown in Figure 7-2.

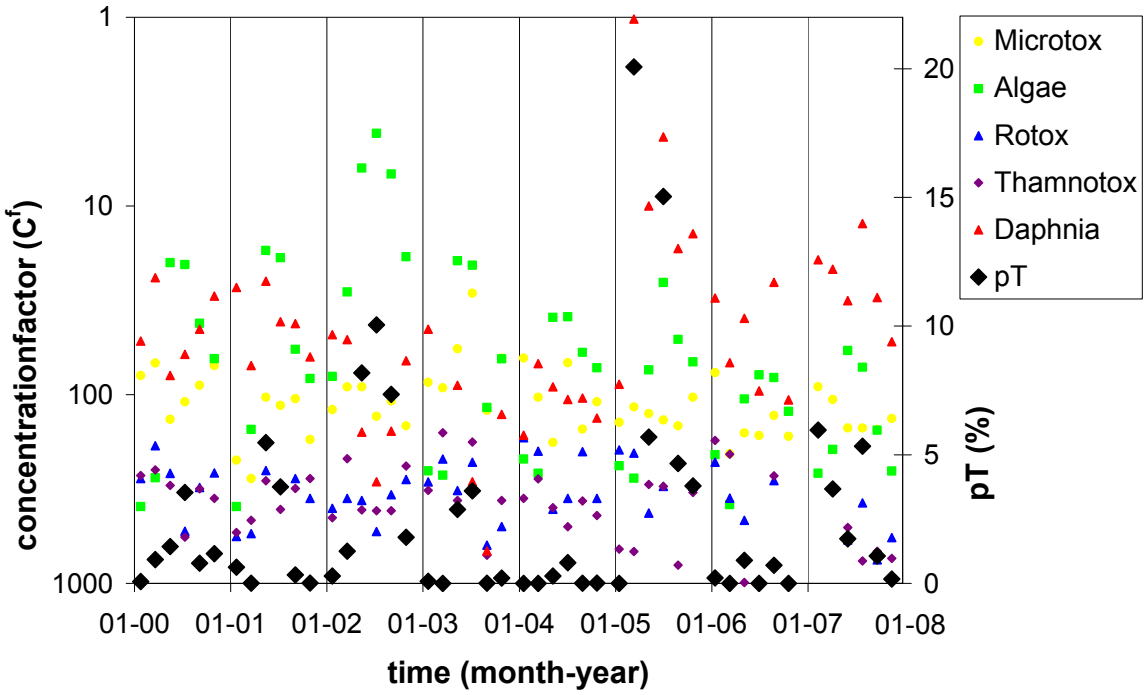


**Figure 7-2. Results of toxicity tests and average pT values for the Meuse at location Eijsden for 2000-2008.**

The Algae PAM test shows most variation within the years. On the contrary, the Microtox® assay and the Thamnotoxkit F™ are more useful to show differences between locations and trends through the years (De Groot et al., 2007).

Since the standard deviation of the pT values calculated for the monitoring data is normally rather large, an assessment method based on relative ranking of toxicity results has been developed (De Groot et al., 2007). All locations were given a rank number, based on the average results for the Microtox®assay and the Thamnotoxkit F™. For both tests, the highest ranks were assigned to location Schaar van Ouden Doel, followed by Eijsden. In Figure 7-3, this toxicity ranking is illustrated for the summarized ranks for Microtox® and Thamnotoxkit F™ together.

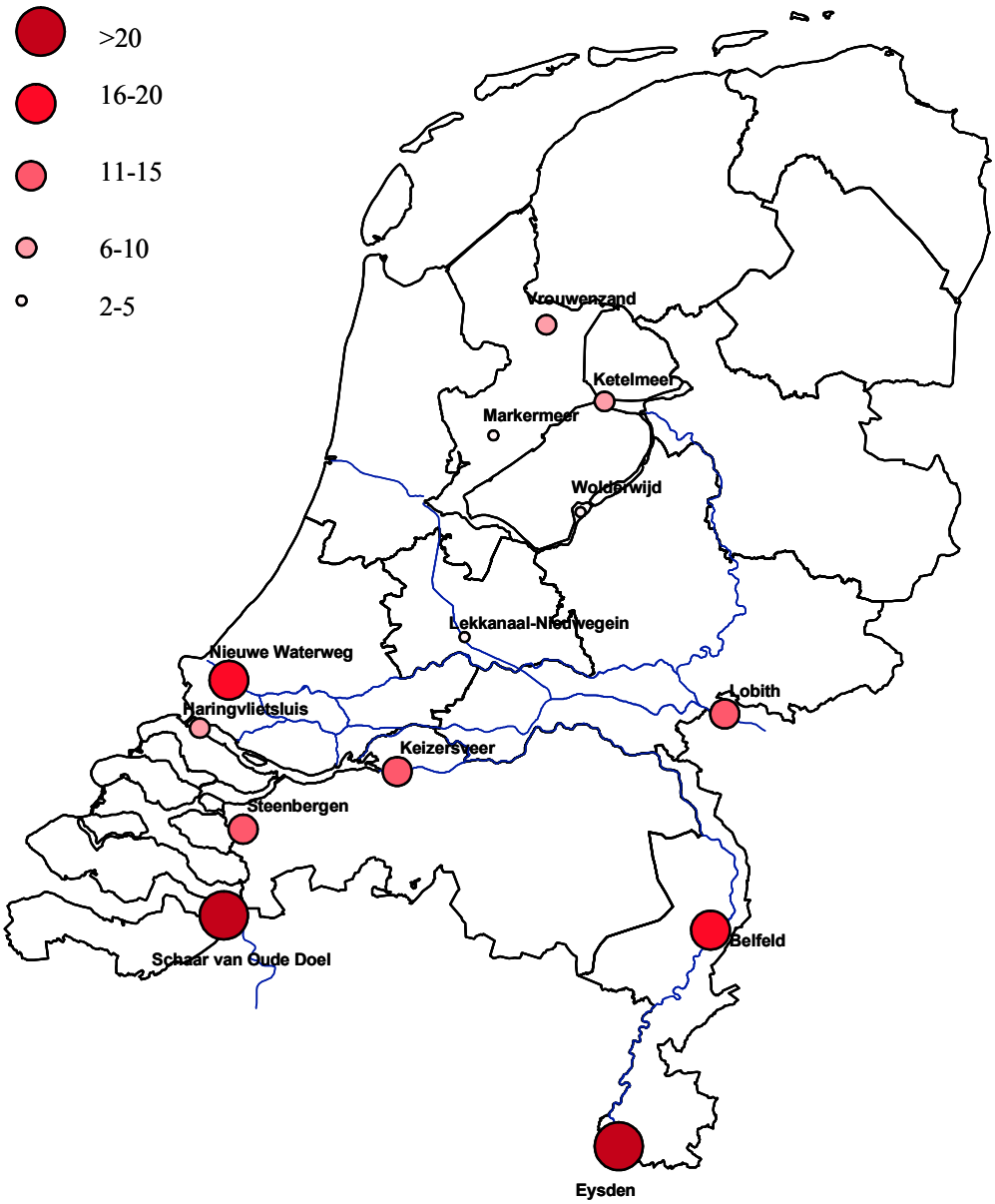


**Figure 7-3. Ranking toxicity (summarized rank numbers) of 13 locations measured from 2000-2008. Ranking of locations was based on average results for the Microtox®assay and the Thamnotoxkit F™, and summarized.**

## 7.3 Identifying sources

A pilot study in the Netherlands was performed to gain information on the toxicity in the canal 'Noordhollands kanaal' by means of bioassays (Roex et al., 2007). Furthermore an attempt to identify the causes of toxicity in the canal was done. Toxicity measurements were performed after XAD-extraction and concentration following the procedures described within this report. Exposure experiments with bacteria, algae and daphnia were carried out to measure acute toxicity. The results of the bioassays were presented as Toxic Units (TU, calculated as $1/C^{f}$. These TU values indicate the

dilution factor for the sample that causes 50 percent acute effect in the assays used. Where the TU value is below 1, the sample needs to be concentrated to cause 50 percent effect.

The first step within the pilot was a screening on the toxicity levels in the canal at the following four locations: Koedijk, Stolpen ‘t Zand and Kooy. These locations were chosen to obtain information on the toxicity levels in the canal downstream the discharge of several municipal wastewater treatment plants and the effect of land use. The agricultural use of the land surrounding the canal changed from grassland to bulb cultivation between the location Koedijk and Stolpen. The use of pesticides within the Dutch flower-bulb culture is high.

The wastewater treatment plants were situated just downstream from the location Koedijk and Stolpen. Figure 7-4 shows the toxicity levels found within this fist screening on toxicity.

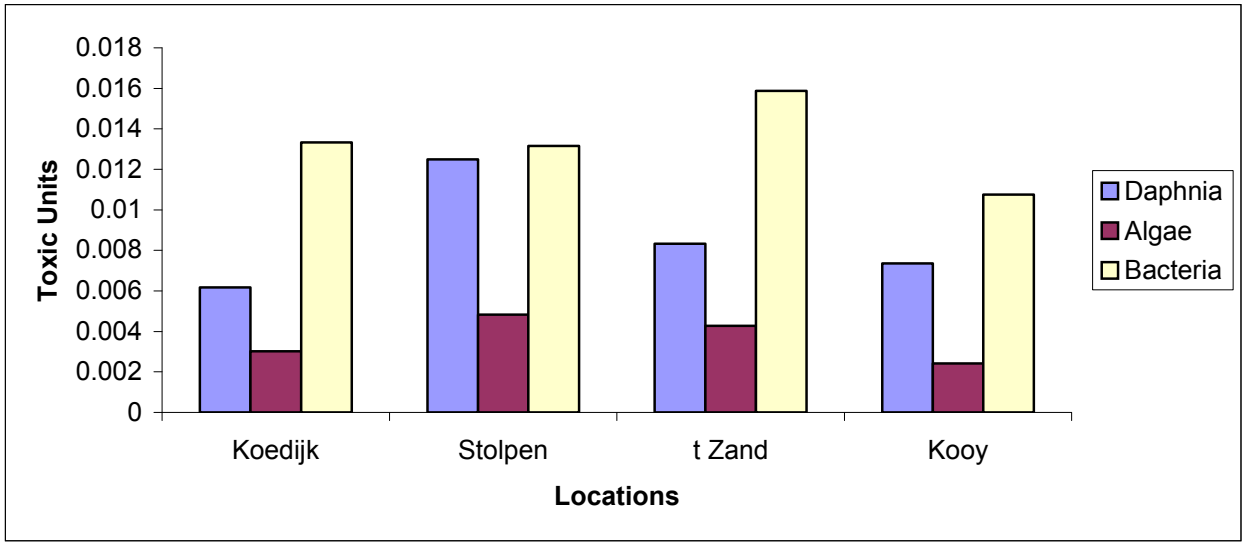


**Figure 7-4. Measured toxicity, expressed in Toxic Units (1/ECf50), on bacteria, algae and Daphnia at several locations in the canal ‘Noordhollands kanaal’.**

An increase in the toxicity on Daphnia and algae can be seen between the locations Koedijk and Stolpen. The effect decreased at the locations ‘t Zand and Kooy. For bacteria an increase in toxicity can be seen at the location ‘t Zand.

In a second sampling campaign the increase in toxicity for Daphnia between Koedijk and Stolpen was investigated. Therefore, five locations between Koedijk and Stolpen were sampled and investigated on toxicity with Daphnia. Daphnia was used for these measurements because this organism had the highest response in the first screening on toxicity. The selected locations were located upstream and downstream from the wastewater treatment plants in this area. Furthermore the effluent from the plants was also tested on toxicity using the same procedure. Figure 7-5 shows the results of the measurements carried out.

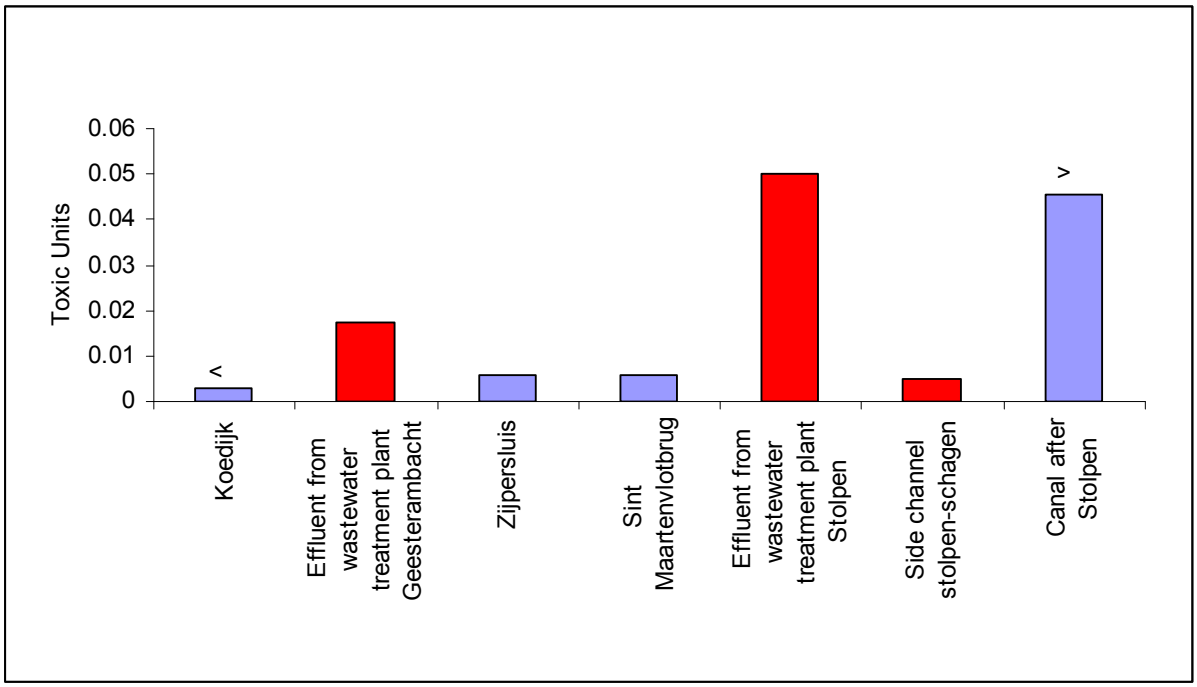


**Figure 7-5. Toxicity on Daphnia after exposure to concentrated samples from locations between Koedijk and Stolpen. The toxicity of the effluent streams from two municipal wastewater treatment plants and a side channel are shown in red. 'Koedijk' is the most upstream location (<) and 'Canal after Stolpen' the most downstream (>).**

The results show that the toxicity on Daphnia from water sampled at location Koedijk is rather low. After the effluent from the first wastewater treatments plant (Geestmerambacht) is discharged into the canal the toxicity slightly increases. The land use (flower and bulb cultivation) between the location Koedijk and Stolpen did not directly increase the toxicity levels found at the locations Zijpersluis and Sint Maartenvlotbrug. After the effluent discharge from the municipal wastewater treatment plant at Stolpen the toxicity of the surface water in the canal increases severely. A water sample from a side channel revealed that the measured toxicity did not originate from this canal. Furthermore the toxicity measured in the effluent stream from the waste water treatment plant at this location was high. These results proof that the effluent discharge ad the location Stolpen is responsible for the toxicity found in the canal. The toxicity encountered after the effluent discharge, approaches the indication of Chronic Effect level (iCE). This should be seen as an indication that the aquatic ecosystem in the canal probably endures a harmful effect caused by the effluent discharge.

## 7.4 Effect monitoring in water management

In 2007 the XAD-method has been applied on several agricultural areas. The aim of this program was to enlarge the knowledge of using effect monitoring with bioassays to water authorities and to examine the surplus value of bioassays in the operational monitoring of pesticides. The program was performed partly in the same time as the chemical monitoring program of the WFD in these areas.

On nine different locations surface water samples were taken together with chemical monitoring of the WFD. These areas varied from an agricultural area, fruit growing culture, cattle breeding and bulb growing industries. References were taken in forest area. Some of the monitoring locations were official WFD locations for WFD chemical monitoring. Other locations were chosen within the water body itself.

Three bioassays were performed using the XAD concentration method: algae growth test, daphnia (conventional 48 h test) and the Microtox®-test. The locations were examined four times during summer and autumn. The bioassays showed severe effects in the concentrated samples. Those effects were not always attributed to the insecticides or herbicides measured in the area. Effects were shown at the indication level of TU=0.05 ($EC^{f}_{50}$ of 20) a few times, which indicates that chronic effect in the original water sample might be achieved (taking a safety factor of 0.5 into account).

This program has shown the use of toxicity assessment in surface water for water authorities. The assessment of toxicity in surface water could be well performed and delivered useful information for water authorities about the use and risk of the applied pesticides in their area.

Methods presented in this report are offered by several commercial parties. Regional water authorities can therefore use the presented techniques for water management themselves. It can help to explain whether or not organic substances can be a fact of concern to reach ecological standards in the water body.

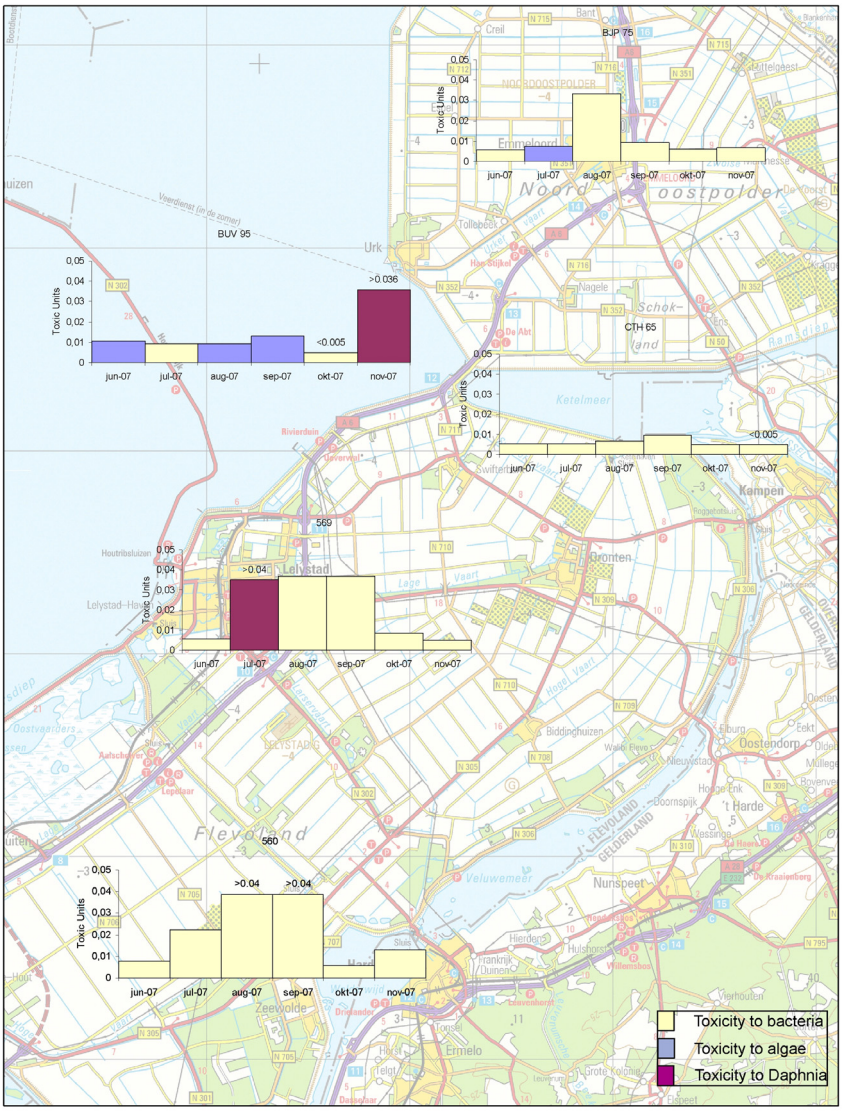


**Figure 7-6. Maximum effect values shown in surface water from Flevoland (Waterboard Zuiderzeeland).**

.

# 8 Discussion

## 8.1 Introduction

In this chapter, the relevance of the method is discussed. The possibilities to improve the detection of additional effects by measuring additional endpoints and the ecological relevance of the method is discussed in section 8.2. In section 8.3, the possibilities for optimisation and the approach for further validation are evaluated. The possibilities for applications of the method and possible improvements are presented in section 8.4. This chapter is closed off with the conclusion in section 8.5, where the possibilities of the method are summarized.

## 8.2 General

### 8.2.1 Detection additional effects by measuring additional endpoints

Bioassays offer a view of the toxicity of substances in water concentrated by the used concentration procedure. The procedure presented in this report works rather good for most persistent apolar organic substances (log $K_{ow}$ > 2). This group of substances include many potential toxic substances for the aquatic ecosystem. Despite of this still limited part of the chemical spectrum that is taken into account, bioassays give a better insight in toxicity than chemical analysis of individual substances (Vonk, 2007), because effects are measured instead of concentrations, combined effects of substances are taken into account and unknown substances are taken into account as well. Thus, compared to the limited set of chemicals measured in surface water, bioassays give more information on whether substances can cause a deviation from the desired ecological quality.

More attention is necessary for substances for which the presented concentration procedure is not optimal. Research focussing on the limitations and possibilities of the concentration procedure for other substances is ongoing. Possibilities for further optimisation of the concentration procedure will be discussed in section 8.2.1. Optimisation of the concentration procedure is especially useful for substance groups which can cause effects in bioassays after a thousand-fold concentration.

Some bioassays can be very specific for certain types of substances while other bioassays reflect toxicity for a broad spectrum of chemical substances. The choice for a certain bioassay depends on the focus of the study: from substance specific to the total toxicity of all substances in surface water. The selected organisms should be representative for organisms in the ecosystem. If a certain substance group is point of interest, a (set) bioassay(s) specifically sensitive to these substances will be sufficient (see also section 7.4).

The appearance of specific groups of substances in surface water, like endocrine disruptors, antibiotics or other pharmaceuticals, can have specific effects on the aquatic ecosystem. At the moment, the impact of these substances is unknown. Specific bioassays can give insight on the impact of these kind of substances on specific locations (Kools et al., 2008). In case of endocrine disruptors, a specific extraction method is needed. In relation to some these specific groups of substances, the explicability of the results and the performance of the concentration procedure (or another extraction method) need further attention.

Preliminary experiments suggest that replacement of the Thamnotoxkit F by the Chydotox-assay would increase the sensitivity of the method (unpublished data RIVM). Disadvantage of such a replacement would be that the time series is interrupted because data before the replacement can not be compared with newer data.

### 8.2.2 Ecological relevance

Within the selection of bioassays for the test battery, ecological relevance is taken into account as good as possible. Until now, it is possible to choose bioassays from three different trophic levels: primary producers represented by an algae test, primary consumers represented by species like the water flea Daphnia and reducers (decomposers) represented by the bacteria *Vibrio fischeri*. Although tests exist for secondary consumers like fish, it is practically impossible to add a fish test due to the sample volume requested for such tests (see also section 4.1). Also, legislation on laboratory animals is a complicating factor. Test duration and ethical objections are additional arguments not to add a fish test to the test battery. The ecological relevance of the test battery could improve if it is possible to add a test which is representative for fish species and can be performed in small sample volumes. The test with the bacteria *Vibrio fischeri* can, to certain extend, be considered representative for fish because this test is representative for the citric acid cycle (Krebs cycle). Cronin et al. (1991) found a strong correlation between data from Microtox® bioassays and fish data.

Although the test battery might not cover all essential trophic levels, it still gives information on the trophic levels for which representative tests are available. If effects are found for a certain test, it can be considered as a signal that effects to the corresponding organism group in the field might occur. Although no efforts have been done for ecological validation yet, the idea is that there is a toxic risk when effects occur in the tests below $C^f$=10 (see section 5.2). An obvious example is the toxicity for algae measured in Eijsden in 2000, which has a clear relationship with the concentrations of three herbicides (Maas et al., 2005). Although there was a suspicion of effects from peak concentrations of insecticides in this study as well, toxicological effects on *Daphnia magna* could not be distinguished from effects of food quality or quantity. For insects, habitat quality seems to play an important role as well.

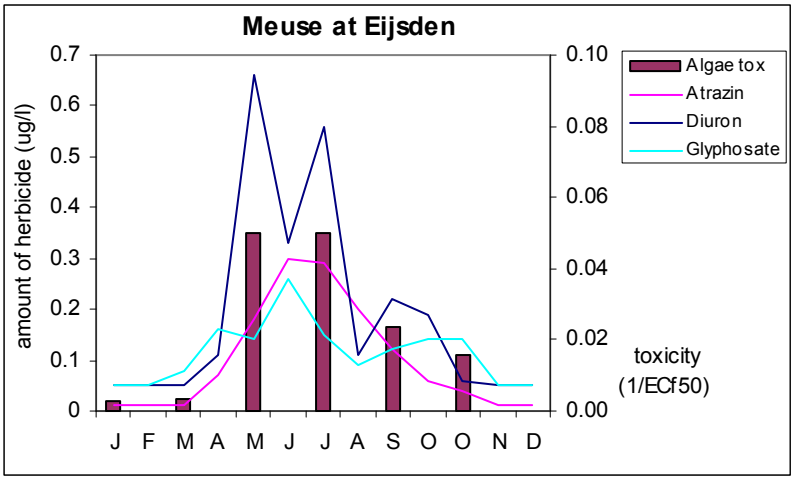


**Figure 8-1. Toxicity for Algae at location Eijsden, measured every two months, and relation to herbicide concentration. Toxicity is presented as Toxic Units (TU=1/ECf50). From Maas et al. (2005).**

Ecological validation will be very complicated because toxicity is only one possible stressor influencing aquatic ecosystems. At the moment, especially factors like a more favourable hydromorphology and less nutrients are expected to contribute obviously to an improvement of Dutch aquatic ecosystems (Ministerie van Verkeer en Waterstaat, 2006). Peeters (2001) compared ecological and ecotoxicological data with physical and chemical measurements. The variation in macrofauna could be explained by habitat quality factors like depth and organic carbon for about 17 %. Toxic substances could explain about 14 % of variation in benthic invertebrate fauna, although the variation could be explained by toxic substances up to 30 % in the large rivers. About 54 % of the variation in benthic invertebrate fauna could not statistically be explained by any of the factors that were taken into account. De Zwart (2005) also found considerable percentages of unexplained variation in comparing fish communities with several stressors like toxic substances, effluents and habitat quality factors. These kind of studies require extensive datasets, which are not available for the monitoring data presented in this report (section 7.2).

On the other hand, the method presented in this report gives a tool to find out whether toxicity might be a cause not to reach ecosystem goals. This is a very useful application for Investigative Monitoring within the Water Framework Directive. If severe toxicity has been found, there is a reason to consider toxicity an important stress factor. Finding toxicity is an obvious reason to search for toxic substances and take measures to reduce toxicity. Although a low toxicity indicates that toxicity does probably not influence the ecosystem quality negatively, the method can not be used to exclude toxicity as a cause. To exclude toxicity as a cause bioassay species should cover more trophic levels or functional groups. Also, the concentrated sample does not contain all toxic substances from the original sample and the substances ratio is not identical to the original sample either. However, bioassays still give more information on whether toxic substances can be a stress factor for the desired ecosystem quality than a limited set of chemical substances would give for which concentrations are measured.

Bioassays can predict the probability that toxic pressure plays a role in reaching the desired ecosystem quality, but can not predict whether species are present or not in a certain ecosystem. Bioassays can not give insight in interactions between species (predation and competition) or cascade-effects, so these interactions should be investigated in the field. Analysis of biodiversity data is always very difficult because of a large natural variation and background error (Vonk, 2007).

## 8.3 Validation

### 8.3.1 Optimisation concentration procedure

Optimisation of the concentration procedure was based mainly on a mixture of narcotic substances. It appeared that the procedure also worked well for several pesticides. Only recently, the possibilities and limitations for some other type of substances were investigated more thoroughly. This shift of interest is prompted by the shift to other problematic substances in water quality management. Because these substances often have different properties than the narcotic substances the concentration procedure was optimised for, the performance of the procedure might be less effective for these new substances of interest. Whether it is useful to optimise the concentration method for new substances of interest will be determined by the usefulness of the following toxicity tests (or other tests suitable for the limited volume resulting from the concentration procedure) in indicating effects from these new substances of interest. If a thousand-fold concentration is not enough to induce acute toxicity for a certain type substance, it might be not useful to optimise the concentration procedure for this kind of substances. For specific substance groups, other extraction methods might be more useful.

In planning monitoring activities using the described procedure and making the concentration procedure available for other organisations, more questions arose on the practical application of the method. In the original time-table for the monitoring, quite some work had to be done during the weekends. With an extended extraction period and/or drying period for the XAD, weekend shifts can largely be avoided which makes the performance of the method more attractive for commercial parties.

Changes in the method have consequences for long-term monitoring, because trends will be broken. To be able to compare with former years after a change of method it is needed to gain insight in the effect. For specific investigations the method can be optimized case by case. Specifically for antibiotics, the extraction might improve by using Amberlite XAD 16 HP, based on the information in the fact sheet of this product (Sigma Aldrich, 1998). This is being investigated at the moment. Some promising biological tests for measuring effects of antibiotics have already been evaluated recently (Kools et al, 2008). The possibilities to measure estrogenic activity have been evaluated in the same study, but this kind of substances certainly needs another extraction method than the method presented in this report.

### 8.3.2 Approach further validation

In chapter 6 a number of performance criteria for analytical methods have been discussed. It is difficult to reach the same approach for the XAD-method as for chemical analytical methods. An important difference between these two methods is that the analytical method is specific for a target concentration (for instance the EQS) and the XAD-method is meant for a range of different mixtures of substances with unknown composition.

The XAD-method and the bioassays have been extensively validated separately (respectively in chapter 3 and 4). A pragmatic approach is followed to estimate the performance criteria for the entire method (chapter 6). The Coefficient of Variation of the bioassay responses and the pT-value seem to have an acceptable range, only the number of results is small.

Since 2007 two more laboratories have been approached to implement the XAD-method. This will lead to a better validation and the opportunity to perform the method more often. The first indication of an inter laboratory survey promises good results. The implementation at these laboratories has to be developed more. This also depends on the use of the method for monitoring purposes. Other studies are already in process. A technical mixture of known substances and known concentrations is made and divided over the three labs. After the XAD-procedure the concentrations of the mixture will be analysed in the extracts (Roex et al., in prep.).

Further validation could be performed to have an indication of the measurement uncertainty at the level of the indication value. Some experimental work has to be fulfilled to make a mixture of substances that will give the proper responses at indication level.

### 8.3.3 Interpretation of the results

Giving an indication on the level of toxicity based on single species tests, is quite transparent compared to the use of species sensitivity distributions. Considering the broadly accepted acute-chronic toxicity ratio of 10 and a combined correction factor of 0.5 for reduced recovery and substances not concentrated, a concentration factor ($C^f$) of 20 should trigger a water manager to take into account that substances might pose an effect on organisms. When all test results show concentration factors exceeding 200, chronic effects will probably not occur in the original sample.

The lack of transparency using statistical extrapolation techniques compared to the use of assessment based on single species is an important drawback that has been put forward more often (EC, 2003). Also, the calculated pT-values for monitoring data presented in this report (section 7.2) are very low for most cases, which makes comparison of locations very difficult. Ranking of bioassays or using average or minimal values offer more possibilities to show differences between locations in these monitoring data (section 7.2).

Exceedance of a certain level gives a very absolute indication of toxicity. In practice, some samples exceed the specified level stronger than others. With presenting a pT value it can be prioritised for which locations substances might pose most effect on organisms. The problem is that the uncertainty borders of the pT values are quite broad. Nevertheless, the probability that the pT level is close to the (presented) mean level is much higher than the probability that the 'true' toxic pressure is close to one of the uncertainty envelopes (Aldenberg and Jaworska, 2000). The mean pT level can be considered as the best fit for the 'true' toxic pressure. The use of statistical extrapolation techniques also provides the possibility to present the indication of toxicity as a probability of exceedance (using a Bayesian approach). Probabilities of exceedance of effect levels give a more detailed view on problematic locations and provide the possibility to prioritize which locations should be dealt with first. Another improvement would be to have a possibility to calculate the Goodness of Fit for the SSD. For a limited number of cases (<10) there are no standard methods to test for Goodness of Fit.

As pT-values are usually quite low for the set of bioassays used in monitoring, ranking of locations based on relative bioassay results is very useful as well to prioritise locations of interest (see section 7.2). Average or minimum values per test might be useful to identify trends in monitoring results for bioassays which are sensitive or insensitive to sampling period respectively. Microtox® and Thamnotoxkit F™ toxicity show the least variation in monitoring within the years, but for algae the results depend on the sampling period within a year (De Groot et al., 2007).

## 8.4 Evaluation of applications of the method

### 8.4.1 Toxic pressure versus toxic potency

Toxicity measured with the chosen bioassays in the test battery gives us an idea of the toxic pressure for the organism groups represented by the chosen tests. Whether a species sensitivity distribution fit through the toxicity measurements (toxic pressure) is representative for an ecosystem as a whole (toxic potency), can be discussed. This discussion is strongly related to the discussion upon ecological relevance (section 8.1.2).

Compared to the criteria on using SSDs for effects assessment in the Technical Guidance Document (TGD) on Risk Assessment (EC, 2003), the data obtained by the available tests cannot be sufficient because they do not match general principles on input data (NOECs prescribed in TGD instead of $EC^f_{50}$/$LC^f_{50}$ data for the presented method), taxonomic groups (3/4 for the presented method instead of 8 taxa prescribed in the TGD) and minimal sample size (maximum of 6 for the presented method instead of mentioned minimum of 10precribed in the TGD). Application of SSDs on toxicity measurements can nevertheless be very useful, because they give a good measure for comparing toxic pressure from different locations as long as the same tests have been done on samples from different locations and the same statistical method has been used to fit the SSD. High values for toxic pressure directly indicate that bioassay organisms will experience effects due to toxicants if they would be present in the ecosystem. This suggests that other organisms likely will experience the same effects in

their natural environment. However, toxic pressure should be interpreted as a relative measure: the higher the toxic pressure, the higher the predicted effects on the tested organisms, the higher the possible effects on the desired ecosystem. A high toxic pressure (certainly compared to other locations) should be a reason to take measures to decrease this pressure (Vonk, 2007).

It is probably too risky to compare calculated toxic pressure based on toxicity measurements with toxic potency for measured concentrations, due to differences in input data, represented taxonomic groups and sample size. It is safer to interpret the values for toxic pressure and toxic potency separately. It might be better to consider comparable input data, taxonomic groups and sample size when it is requested to compare toxic pressure for extracted toxicants with toxic potency for specific chemically measured chemicals.

Analogous to toxic pressure, a high toxic potency for specific chemically measured chemicals for a certain location compared to other locations can be a reason to take measures to decrease this pressure as well.

### 8.4.2 Usefulness for WFD

Bioassays have no official status as an instrument within the Water Framework Directive, because the WFD is substance-oriented instead of effect-oriented. At the moment, bioassays cannot replace any (part of) the obligated chemical analysis within the WFD. An effect found in bioassays does not give insight in which substances cause this effect and which measures could be taken to prevent this effect. Nevertheless, bioassays can be used complementary to the chemical monitoring as a first screening step, which will give more information on total toxic pressure instead of concentrations of priority pollutants alone. Although monitoring costs will increase, this additional information can eventually lead to cost reduction because the budget for measures can be used more efficiently.

Bioassays can be used within WFD monitoring:

1. for additional information on the condition of a water system. This gives insight in the total toxic pressure instead of concentrations of priority substances alone;
2. to compare locations in space and time to assess whether locations are representative for monitoring. This can keep the number of locations within the full WFD monitoring program restricted;
3. to demonstrate whether toxic substances contribute to effects on ecology. Bioassay results can be used to confirm and explain measures to solve chemical pressure in relation to other pressure factors;
4. as a monitoring instrument to monitor the effect of measures , or prioritise locations to take measures;
5. to prioritise the most important substances or substance groups by using substance(group)-specific bioassays or biomarkers.

The most promising application for bioassays within WFD is diagnosis (option 3 mentioned above), because it doesn't imply additional costs for this purpose. This gives bioassays a role within the investigative monitoring part of the WFD and is therefore an issue in the Dutch 'Diagnostic Water Quality Instrument' (Van Riel and Knoben, 2007; Maas et al., 2008). When a high toxicity is found in bioassays, toxicity can be an important stress factor. When only a low toxicity is found with the pT-method, it is (with the present knowledge) quite sure that toxicity is not a (major) negative factor for the ecological condition of a water system. This is not 100 % for sure, because not all substances are taken into account. However, compared to the limited set of chemical substances for which concentrations are analysed, bioassays give more confidence whether substances are expected to be a

major factor for deviation of the desired ecological quality. This does not imply that a good ecological status will be reached with reducing toxicity. Ecological quality can be influenced by many stressors, like pH, eutrophication or hydromorphology (Heugens et al., 2001). The eco-epidemiological approach is a promising technique to discriminate the role of toxicants from other potential stressors (De Zwart, 2005).

Although bioassays can be useful additionally to surveillance monitoring, it is expected that it won't be used that way within short notice within the WFD. For the first period until 2015, monitoring programs have already been made and additional research will only increase monitoring costs. Within operational monitoring, bioassays can be useful to monitor the biological effectiveness of measures. Operational monitoring is used for locations where problematic substances are known, which have to be monitored chemically. Bioassays can only be used additional to the chemical monitoring, so in that case the surplus value will be small.

### 8.4.3 WEA application

Whole Effluent Assessment is an effect-based method to judge the environmental effect of effluents. As presented in section 7.3, this method can be used to identify the sources for toxic effects. It can also be used to investigate to what extend an effluent causes toxic effects within a water system. Which bioassays should be used and how the sample should be concentrated depends on the type of substances expected. More specific concentrations techniques and bioassays can be used for specific substance groups, and can be interpreted with simpler techniques. The presented bioassays in this report do not cover all toxic modes of action. The concentration technique presented in this report works rather well for narcotic substances, but can have restrictions for other type of substances. Especially for upcoming substances, new techniques are necessary and are studied at the moment (e.g. antibiotics).

## 8.5 Conclusions

- The presented method can partially provide information on effects of substances on organisms;
- For specific purposes, a simple setup with a limited set of bioassays and a simple interpretation method gives sufficient information on problems and their causes in water systems;
- For an ecologically relevant approach, the method should be extended with other representative test organisms and more sensitive methods.

The presented method is useful for monitoring goals and setting priorities in water quality management, although it should be taken into account that the presented method is not suitable for all type of substances possibly present in the ecosystem. Investments in new extraction methods are useful, especially for more polar and new substances, because these types of substances are increasingly relevant for the environment.

Demonstrating relevant ecological risks will especially be useful for WFD purposes. Water managers will be able to demonstrate whether the presence of toxic substances influences the Good Ecological Status and to what extend toxic substances will harm the ecosystem.

In addition, the presented method is suitable for diagnosis and is a solid ground for considering measures.

rivm

# List of abbreviations

| | |
|---|---|
| ACR | Acute-Chronic toxicity Ratio |
| BAT | Best Available Technique |
| BLM | Biotic Ligand Model |
| Cf | Concentration factor |
| CDF | Cumulative Distribution Function |
| EC50 | median Effective Concentration |
| ECf50 | median Effective Concentration factor |
| ELV | Emission Limit Values |
| EQS | Environmental Quality Standard |
| ERL | Environmental Risk Limit |
| FA | Fraction Affected |
| HC5 | Hazardous Concentration to 5 % of the species |
| HC50 | Hazardous Concentration to 50 % of the species |
| iCE | concentration level indicating Chronic Effect |
| iNE | concentration level indicating Negligible Effect |
| IPPC | Integrated Pollution Prevention and Control |
| LC50 | median Lethal Concentration |
| LCf50 | median Lethal Concentration factor |
| LOD | Limit Of Detection |
| LOQ | Limit Of Quantification |
| PAF | Potentially Affected Fraction |
| PBT | Persistence, Bioaccumulation and Toxicity |
| pT | Toxic pressure |
| QA | Quality Assurance |
| QC | Quality Control |
| RIVM | National Institute for Public Health and the Environment, abbreviation for name in Dutch: Rijksinstituut voor Volksgezondheid en Milieu |
| RIZA | Institute for Inland Water Management and Waste Water Treatment, abbreviation for name in Dutch: Rijksinstituut voor Integraal Zoetwaterbeheer en Afvalwater. Currently: Centre of Water Management (since 1-10-2008) |
| SSD | Species Sensitivity Distributions |
| TGD | Technical Guidance Document |
| TU | Toxic Unit |
| U | measurement Uncertainty |
| WEA | Whole Effluent Assessment |
| WFD | Water Framework Directive |
| XAD | synthetic resin |

rivm

# References


Aiken GR, Thurman EM, Malcolm RL (1979). Comparison of XAD macroporous resins for the concentration of fulvic-acid from aqueous solution. Anal. Chem., 51, 11, 1799-1803.

Aldenberg T, Jaworska JS (2000). Uncertainty of the Hazardous Concentration and Fraction Affected for Normal Species Sensitivity Distributions. Ecotoxicology and Environmental Safety 46, 1-18.

Aldenberg T, Jaworska JS, Traas TP (2002). Normal Species Sensitivity Distributions and Probabilistic Ecological Risk Assessment. Published in: Species Sensitivity Distributions in Ecotoxicology. Posthuma L, Suter II GW, Traas, TP (eds.). Lewis Publishers, Boca Raton, FL, pp 49-102.

Aqua Survey, Inc. (1997). Daphnia IQ toxicity test kit™. Daphnia toxicity screening test for freshwater, instructions. Aqua Survey, Inc., 499 Point Breeze Rd., Flemington, NJ, USA.

AWWA-KIWA (1988). The search for a surrogate, cooperative research report, AWWA research foundation, American Water Works Association, ISBN 0-89867-438-7.

Blaise C, Legault R, Bermingham N, Van Collie R, Vasseur P (1986). A simple microplate algal assay for aquatic toxicity assessment. Tox. Assess. Int. Q, 1, 261-281.

Bulich AA (1979). Use of luminescent bacteria for determining toxicity in aquatic environments. In: Marking LL and Kimerle RA (eds.). Aquatic Toxicology. STP 667. American Society for Testing and Materials, Philadelphia, PA, pp. 98–106.

Bulich AA and Isenberg DL (1981). Use of luminescent bacterial system for the rapid assessment of aquatic toxicity. ISA Trans 20, pp. 29–33.

Burke, J (1984). Solubility Parameters: Theory and Application. Published in: AIC Book and Paper Group Annual, Volume 3, Ed.: Jensen C. Page 13-58.

Calow P (1993). Handbook of Ecotoxicology, Vol 1: Blackwell Scientific Publications, Oxford, England.

Centeno MDF, Persoone G, Goyvaerts MP (1995). Cyst-based toxicity tests IX. The potential of *Thamnocephalus platyurus* as test species in comparison with *Stretocephalus proboscideus* (Crustacea: Branchipoda: Anostraca). Environmental Toxicology and Water Quality.10 (4), 275-282.

Collombon MT (2007). Concentration of 'forgotten' substances using the XAD concentration method. Suitability of the method for hydrophilic chemicals. RIVM letter report 607013001.

Collombon M, Van de Kamp R, Struijs J (1997). Procedures for extracting organic micro-pollutants from water samples to monitor toxicological stress. RIVM report 607042008.

Creasel (1990). Rotoxkit F™ Rotifer Toxicity Screening test for freshwater. Standard Operational Procedure V071090. Creasel Ltd., Deinze, Belgium.

Creasel (1992). Thamnotoxkit F™. Crustacean Toxicity Screening test for freshwater. Standard Operational Procedure V241092. Creasel Ltd. Deinze, Belgium.

Cronin MTD, Dearden JC, Dobbs AJ (1991). QSAR studies of comparative toxicity in aquatic organisms. Science of the total environment 109/110: 431-439.

De Groot A, Verweij W, en Maas J (2003). Toxicologische risico's van de Rijn en de Maas. $H_2O$, 25/26; 31-33 (in Dutch).

De Groot AC, Maas JL, Wouterse MJ, Espeldoorn A, Van Dijk B (2006). De Potentiële toxiciteit van oppervlaktewater uit Bovensluis, Eijsden, Haringvlietsluis, Ketelmeer, Lobith, Steenbergen, Schaar van Ouden Doel en Wolderwijd. LER-notitie 02/06. RIVM-LER (in Dutch).

De Groot AC, Wouterse MJ, Van Dijk GG, Mulder C (2007). De bepaling van toxische druk in Rijkswateren. RIVM letter report 607013002 (in Dutch).

Dekker T, Greve GD, Ter Laak TL, Boivin ME, Veuger B, Gortzak G, Dumfries S, Lücker SMG, Kraak MHS, Admiraal W, Van der Geest HG (2006). Development and application of a sediment toxicity test using the benthic cladoceran *Chydorus sphaericus*. Environmental Pollution 140, 231-238.

De Nijs ACM, Driesprong A, den Hollander HA, De Poorter LRM, Verweij W, Vonk JA, De Zwart D (2008). Risico's van toxische stoffen in de Nederlandse oppervlaktewateren. RIVM report 607340001 (in Dutch).

Derksen JGM, Van Eijnatten GM, Lahr J, Van der Linde P, Kroon AGM (2001). Milieu-effecten van humane geneesmiddelen, aanwezigheid en risico's. RIZA report 2001.051 (in Dutch).

De Zwart D (2002). Observed Regularities in Species Sensitivity Distributions. Published in: Species Sensitivity Distributions in Ecotoxicology. Posthuma L, Suter II GW, Traas, TP (eds.). Lewis Publishers, Boca Raton, FL, pp 133-154.

De Zwart D, Sterkenburg A (2002).Toxicity-based assessment of water quality. Published in: Species Sensitivity Distributions in Ecotoxicology. Posthuma L, Suter II GW, Traas, TP (eds.). Lewis Publishers, Boca Raton, FL, pp 383-402.

De Zwart D (2005). Impact of Toxicity on Species Composition of Aquatic Communities: Concordance of Predictions and Field Observations. Thesis University of Amsterdam, Faculty of Science.

EC (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Communities L327/1.

EC (2003). Technical Guidance Document on Risk Assessment. In support of Commission Directive 93/67/EEC, Commission Regulation (EC) No. 1488/94 and Directive 98/8/EC. http://europa.eu.int (download 29-11-2006 15:33:52).

Freeman, W (1953). A standardized method for determining toxicity of pure compounds to fish. Sewage and industrial wastes, 25: 845-848.

Genty B, Briantais JM, Bakaer NR (1989). The relationship between quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. Biochim. Biophys. Acta, 990, 87-92.

Griffioen M, Van Weerdum G (2007) Commentaar op artikel 'Watertypespecifieke risicogrenzen en afwenteling' $H_2O$, 40:19 (in Dutch).

Grothe DR, Kimerle RA (1988). Inter- and intralaboratory variability in Daphnia magna effluent toxicity test results. Environ. Toxicol. Chem. 4, 189 – 192: From: Persoone and Janssen; Handbook of Ecotoxicology, Vol 1: P. Calow, Blackwell Scientific Publications, Oxford, England.

Haanstra L, Doelman P, Oude Voshaar JH (1985) The use of sigmoidal dose response curves in soil ecotoxicological research, Plant and Soil 84 (1985), pp. 293–297.

Heugens EHW, Hendriks AJ, Dekker T, Van Straalen NM, Admiraal W (2001) A review of effects of multiple stressors on aquatic organisms and analysis of uncertainty factors for use in risk assessment. Critical reviews in toxicology, 31 (3): 247-284.

Hofstraat JW, Snel JFH, Geel C (1994). Simple determination of photosynthetic efficiency and photoinhibition of *Dunaliella tertiolecta* by saturating pulse fluorescence measurements. Mar. Ecol. Prog. Ser. 103, 187 – 196.

ISO 6431 (1996). Water quality – Determination of the inhibition of the mobility of *Daphnia magna* Straus (Cladocera, Crustacea) – Acute toxicity test. International Organization for Standardization, Geneva, Switzerland. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=21923 (download December 2008).

ISO 11348-3 (1998). Water quality – Determination of the inhibitory effect of water samples on the light emission of Vibrio fischeri (Luminescent bacteria test) – Part 3: Method using freeze-

dried bacteria. International Organization for Standardization, Geneva, Switzerland. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40518 (download December 2008).

ISO 5667-3 (2003). Water quality – Sampling – Part 3: Guidance on the preservation and handling of water samples. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=33486 (download December 2008).

ISO 8692 (2004). Water quality -- Freshwater algal growth inhibition test with unicellular green algae. International Organization for Standardization, Geneva, Switzerland. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=29924 (download December 2008).

ISO 5667-6 (2005). Water quality - Sampling - Part 6: Guidance on sampling of rivers and streams. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34815 (download December 2008).

Janssen CR, Persoone G (1993). Rapid toxicity screening tests for aquatic biota. Methodology and experiments with Daphnia magna. Environmental toxicology and chemistry, vol 12 pp. 711-717.

Kools SAE, Van den Brandhof EJ, Mons MN, Hogenboom AH, Derksen JGM (2008). Verkenning geneesmiddelen en toxiciteit effluent RWZI's. STOWA report 2008-06 (in Dutch).

Lepom P, Hanke G, Wollgast J, Quevauviller P (2008). Guidance on surface water chemical monitoring under the Water Framework Directive. Draft, version 9, 2008.

Luoma SN, Ho K (1993). Appropriate uses of marine and estuarine sediment bioassays. From: Handbook of Ecotoxicology, Vol 1: P. Calow, Blackwell Scientific Publications, Oxford, England.

Maas JL, Van de Plassche E, Straetmans A, Vethaak AD, Belfroid AC (2003). Normstelling voor bioassays. Uitwerking voor oppervlaktewater en waterbodem. RIZA report 2003.005, RIKZ report 2003.007 (in Dutch).

Maas JL (2005). Discussion note (non-status paper). Description of scenarios of bio-analyses in Operational Monitoring of the WFD. RIZA, July 2005, working group 6/7, Swift-WFD.

Maas JL, Greijdanus M, Beek MA (2005). Ecotoxicologische & ecologische monitoring Maas (EMMA-water). RWS-RIZA working document 2005.182X (in Dutch).

Maas JL, Van den Heuvel-Greve MJ (2005).Opportunities for bio-analysis in WFD chemical monitoring using bioassays. RWS-RIZA working document 2005.053X.

Maas JL, Schipper CA, Knoben RAE, Van den Heuvel-Greve MJ, Den Besten PJ, De Maagd PGJ (2008). Diagnostic Water Quality Instruments for the use in the Water Framework Directive. Chapter 3.2 in: The Water Framework Directive – Ecological and chemical status monitoring, Edited by Quevauviller, Borchers, Thompson and Simonart, John Wiley and Sons, Ltd.

Mensink BJWG, Smit CE, Montforts MHMM (2008). Manual for summarising and evaluating environmental aspects of plant protection products, RIVM report 601712004.

Microbics (1992). Microtox® Manual. Microbics Corporation, Carlsbad, CA.

Ministerie van Verkeer en Waterstaat (2006). Decembernota 2006 KRW/WB21 Beleidsbrief. Ministerie van V&W, Ministerie van LNV en het Ministerie van VROM met medewerking van IPO, VNG, Unie van Waterschappen en de Regionale Ambtelijke Overleggen (in Dutch).

OECD 202 (2004). OECD Guideline for Testing of Chemicals. *Daphnia* sp., Acute Immobilisation Test. OECD/ECDE. Adopted 13 April 2004. http://www.oecdbookshop.org/oecd/display.asp?lang=en&sf1=DI&st1=5LMQCR2K7S28 (download December 2008).

Peeters ETHM (2001) Benthic macroinvertebrates and multiple stressors. Quantification of the effects of multiple stressors in field, laboratory and model settings. Thesis Wageningen Universtity, the Netherlands.
Persoone G, Blaise C, Schnell T, Janssen C, Van Steertegem W (1993). Cyst-based toxicity tests: I. Report on an international intercalibration exercise with three cost-effective Toxkits. Z. Angew. Zool., 79, 17-36.
Persoone G (1995). Development and validation of toxkit microbiotests with invertebrates, in particular crustaceans. Published in: Microscale aquatic toxicology, advances, techniques and practice. Wells PG, Lee K and Blaise C (eds.). CRC Lewis Publishers.
Pieters BJ, Bosman-Meijerman D, Steenbergen E, Van den Brandhof EJ, Van Beelen P, Van der Grinten E, Verweij W, Kraak MHS (2008). Ecological quality assessment of Dutch surface waters using a new bioassay with the cladoceran *Chydorus sphaericus*. Proc. Neth. Entomol. Soc. Meet. Vol 19: 157-164.
Posthuma L, Suter II GW, Traas TP ,eds. (2002). Species Sensitivity Distributions in Ecotoxicology. Boca Raton: Lewis Publishers.
Postma JF, De Valk S, Dubbeldam M, Maas JL, Tonkes M, Schipper CA, Kater BJ (2002). Confounding factors in bioassays with freshwater and marine organisms. Ecotoxicology and Environmental Safety 53: 226-237.
Prichard E (2004). Guidelines for Laboratories carrying out Measurements where the results will be used to implement the Water Framework Directive (2000/60/EC). Swift-WFD, D12 Part 1.
Prichard E (2006). Guidelines for laboratories carrying out measurements where the results will be used to implement the Water Framework Directive (2000/60/EC). Swift-WFD, D12 Part 2.
Quevauviller P (2008). Adopting technical specification for chemical monitoring and quality of analytical results in accordance with Directive 2000/60/EC.Draft Commission Decision. EAQC-WISE-project. http://www.eaqc-wise.net/ (download December 2008).
Raimondo S, Montague BJ, Barron MG (2007). Determinants of variability in acute to chronic ratios for aquatic invertebrates and fish. Environmental Toxicology and chemistry, Vol. 26, No. 9, pp 2019-2023.
Ribo JM, Riva MC (2004). 12th inter-laboratory comparison exercise for the luminescent bacteria toxicity assay. Environmental Toxicology Laboratory, Technical University of Catalonia (UPC).
Riva MC, Ribo JM (2006). Inter-laboratorium comparison exercise for the alga and daphnia toxicity test. Environmental Toxicology Laboratory, Technical University of Catalonia (UPC).
Roex E (2005). Nut en noodzaak van het extraheren van effluenten voor de TEB-methodiek. RIZA report 2005.012 (in Dutch).
Roex E, Rotteveel S, Ferdinandy M, Bakker V (2007). Totaal-effluentbeoordeling toegepast in de praktijk. $H_2O$ 10, 55-57 (in Dutch).
Roex E, Rotteveel S (in prep.). De meerwaarde van TEB. Totaal-effluentbeoordeling als instrument voor het prioriteren van bronnen en het beoordelen van de effectiviteit van aanvullende zuiveringsmaatregelen. Growth document, 2008. Deltares, in prep. (in Dutch).
Roex E, Stroomberg G, Van der Grinten E, Rotteveel S (in prep.). Intra- en interlab variatie van XAD-extracties en bioassays in oppervlaktewater. Deltares report, in prep. (in Dutch).
Roig B, Allan IJ, Greenwood R (2003). SWIFT-WFD. Screening methods for Waterdata InFormaTion in support of the implementation of the Water Framework Directive. A Toolbox of existing and emerging methods for water monitoring under the WFD. Report: Draft 2.
Rotteveel S, Roex E (2006). Variatie in TEB-testen voor acute toxiciteit met onbehandelde en geëxtraheerde monsters. RWS-RIZA working document 2006.074X (in Dutch).
RWS (1999). Monitoring Zoete Rijkswateren. Authors: Gilde LJ, Prins KH and Van Helmond CAM. RIZA report 99.004. ISBN 90 369 52 247 (in Dutch).

Sigma-Aldrich (1998) Product Information Amberlite XAD Polymeric Resins. http://www.sigmaaldrich.com/sigma/product%20information%20sheet/xad7pis.pdf. (download December 2008).
Slooff W, Van Kreul CF, De Zwart D (1984). Biologische parameters en oppervlaktewater(meetnetten). $H_2O$ (17)1: 2-5 (in Dutch).
Snell TW, Persoone G (1989). Acute toxicity bioassays using rotifers. II. A freshwater test with *Brachionus rubens*. Aquatic Toxicology, 14, 81 – 92.
Struijs J, Van Buren L (1995). Milieutoxiciteit (pT) van water (I): het opwerken tot waterconcentraten. RIVM report 607042006 (in Dutch).
Struijs J, Van de Kamp R, Hogendoorn EA (1998). Isolating organic micropollutants from water samples by means of XAD resins and supercritical fluid extraction. RIVM report 607602001.
Struijs J, Ritsema R, Van de Kamp R, De Zwart D (2000). Toxic pressure in surface water. A pilot of new monitoring techniques. RIVM report 607200003.
Struijs J, De Groot AC (2001). Letter report project 'Toxische druk metalen in water'. Projectnumber S/607501/01/EB. RIVM-LER (in Dutch).
Struijs J, Van de Kamp RE (2001). A revised procedure to concentrate organic micro-pollutants in water. RIVM report 607501001.
Struijs J, De Zwart D (2003). Evaluatie van pT. De bepaling van toxische druk in water. RIVM report 860703001 (in Dutch).
Suter GW, Traas TP, Posthuma L (2002). Issues and Practices in the Derivation and Use of Species Sensitivity Distributions. Published in: Species Sensitivity Distributions in Ecotoxicology. Posthuma L, Suter II GW, Traas, TP (eds.). Lewis Publishers, Boca Raton, FL, pp 437-474.
Traas TP, ed. (2001). Guidance document on deriving Environmental Risk Limits. RIVM report 601501012.
U.S. EPA, (1985). Methods for measuring the acute toxicity of effluents to freshwater and marine organisms. EPA 600/4-85-013 Environ. Research Laboratory, Duluth.
Vaal MA, Folkerts, AJ (1998). Sensitivity of microscale ecotoxicity tests and their suitability to measure toxicity of environmental samples. RIVM report 607042009.
Vaal MA, Folkerts AJ, Van de Kamp, RE, Struijs J (1999). Toxicological validation of a procedure for extracting organic micropollutants form water samples. RIVM report 607200002.
Van Beusekom SAM, Admiraal W, Sterkenburg A, De Zwart D (1999). Handleiding PAM-Test. ECO notitie 98/09, Laboratorium voor Ecotoxicologie, RIVM (in Dutch)
Van Riel MC, Knoben RAE (2007). Handreiking diagnostiek. Ecologische kwaliteit van watersystemen. Royal Haskoning in opdracht van Rijkswaterstaat. Versie 03. November 2007 (in Dutch).
Van Stee L, Leonards PEG, Van Loon WMGM, Hendriks AJ, Maas JL, Struijs J, Brinkman UAT (2002). Use of semi-permeable membrane devices and solid-phase extraction for the wide-range screening of microcontaminants in surface water by GC-AED/MS. Water Research 36: 4455-4470.
Van Straalen NM, Denneman CAJ (1989). Ecotoxicological evaluation of soil quality criteria. Ecotoxicology and Environmental Safety, 18:241-251.
Van Vlaardingen PLA, Traas TP, Wintersen AM, Aldenberg T (2004). ETX 2.0. A Program to Calculate Hazardous Concentrations and Fraction Affected, Based on Normally Distributed Toxicity Data. RIVM report 601501028.
VIM (1996) International Vocabulary of Basic and General Terms in Metrology. 2nd Edition, ISO, Geneva, 1993, ISBN 92-67-01075-1.
Vonk JA (2007). Relevantie van toxiciteit-bioassays bij de beoordeling van een goede toestand van oppervlaktewater. RIVM letter report 607013006 (in Dutch).

Willemsen A, Vaal MA, De Zwart D (1995). Microbiotests as tools for environmental monitoring. RIVM report 607042005.

Zalewka-Radziwill M (2000). Toxicity assessment of chemicals using conventional acute *Daphnia magna* tests, Toxkits and Fluotox microbiotests. From: New microbiotests for routine toxicity screening and biomonitoring, ed. Persoone et al., 2000. Kluwer Academic/Plenum Publishers, New York.

# Appendix I. Protocols

Inclusion of all protocols in the printed version would mean a large number of pages and a lot of paper. Therefore, the protocols will be included only in the on-line edition of the report, which can be found at www.rivm.nl/bibliotheek/rapporten/607013010.html

I.1 Sampling
I.2 XAD extraction
I.3 Microtox® assay
I.4 Algae: PAM-test
I.5 Algae: Fluorescence test (growth)
I.6 Daphnia IQ test
I.7 Daphnia immibility test
I.8 Thamnotoxkit F™
I.9 Rotoxkit F™
I.10 Chydorid test

rivm

# Appendix II. Details on chemicals: substance properties and CAS-numbers

| Substance type | Substance | CAS nr | mol. formula | water solubility ($mg\ l^{-1}$) 20-25 ºC | log $K_{ow}$ | Henry's law constant ($Pa\ m^3\ mol^{-1}$) | vapour pressure (Pa) 20-25 ºC | boiling point (ºC) |
|---|---|---|---|---|---|---|---|---|
| narcotic | 1,4-dichlorobenzene* | 106-46-7 | $C_6H_4Cl_2$ | 83 | 3.5 | 160 | 90 | 174 |
| narcotic | hexachloroethane | 67-72-1 | $C_2Cl_6$ | 3.1 | 4.6 | 6800 | 90 | 186 |
| narcotic | 1,3,5-trichlorobenzene* | 108-70-3 | $C_6H_3Cl_3$ | 25 | 4.0 | 200 | 28 | 208 |
| narcotic | 3,4-dichlorotoluene | 95-750 | $C_6H_3Cl_2OH$ | 10.5 | 4.2 | 650 | 42 | 208 |
| narcotic | 1,2,3-trichlorobenzene* | 87-61-6 | $C_6H_3Cl_3$ | 17 | 3.8 | 300 | 28 | 218 |
| narcotic | 3-chloronitrobenzene | 121-73-3 | $C_6H_4ClNO_2$ | 501 | 2.6 | 0.09 | 0.3 | 235 |
| narcotic | 2,4-dichloroaniline | 554-00-7 | $C_6H_5Cl_2N$ | 214 | 2.8 | 0.67 | 0.9 | 245 |
| narcotic | 1,2,3,4-tetrachlorobenzene | 634-662 | $C_6H_2Cl_4$ | 4.3 | 4.3 | 250 | 0.7 | 216 |
| narcotic | 3,4-dichloro-nitrobenzene | 99-54-7 | $C_6H_3Cl_2NO_2$ | 97 | 3.3 | 0.084 | 0.04 | 255 |
| narcotic | 2,4,6-trichloroaniline | 634-93-5 | $C_6H_4Cl_3N$ | 26 | 3.6 | 2.2 | 0.3 | 262 |
| narcotic | pentachlorobenzene* | 608-93-5 | $C_6HCl_5$ | 0.6 | 4.8 | 980 | 2.2 | 277 |
| pesticide | mevinphos | 7786-34-7 | $C_7H_{13}O_6P$ | >1000 | 0.1 | $< 3.8 \times 10^{-3}$ | $1.7 \times 10^{-2}$ | |
| pesticide | lindane** | 58-89-9 | $C_6H_6C_{l6}$ | 12.7 | 3.8 | $1.5 \times 10^{-2}$ | $5.6 \times 10^{-4}$ | 323 |
| pesticide | diazinon | 333-41-5 | $C_{12}H_{21}O_3N_2PS$ | 60 | 3.3 | $6.1 \times 10^{-2}$ | $1.2 \times 10^{-2}$ | |
| pesticide | parathion-methyl | 298-00-0 | $C_8H_{10}NO_5PS$ | 55 | 3.4 | $9.8 \times 10^{-4}$ | $2 \times 10^{-4}$ | 154 |
| pesticide | fenchlorfos | 299-84-3 | $C_8H_8C_{l3}O_3PS$ | 1.1 | 5.3 | 29 | $1 \times 10^{-1}$ | |
| pesticide | chlorfenvinfos* | 470-90-1 | $C_{12}H_{14}C_{l3}O_4P$ | 95 | 3.2 | $3.8 \times 10^{-4}$ | $1 \times 10^{-4}$ | 167-170 |

* Priority Substance, according to Water Frame Directive, Daughter directive, proposal 17 July 2006;
** Priority Hazardous Substance, according to Water Frame Directive, Daughter directive, proposal 17 July 2006;
All these chemicals can be analysed by GC.

| Substance type | Substance | CAS nr | mol.formula | water solubility ($mg\ l^{-1}$) 20-25 ºC | log $K_{ow}$ | Henry's law constant ($Pa\ m^3\ mol^{-1}$) | vapour pressure(Pa) 20-25 ºC | Analysis |
|---|---|---|---|---|---|---|---|---|
| pesticide | metoxuron*** | 19937-59-8 | $C_{10}H_{13}O_2NCl$ | 678 | 1.6 | $1.45 \times 10^{-3}$ | $4.3 \times 10^{-3}$ | HPLC-UV |
| pesticide | diuron* | 330-54-1 | $C_9H_{10}Cl_2ON_2$ | 42 | 2.8 | $6.1 \times 10^{-6}$ | $1.6 \times 10^{-6}$ | HPLC-UV |
| pesticide | azinphos-methyl | 86-50-0 | $C_{10}H_{12}O_3N_3PS$ | 30 | 3.0 | $1.9 \times 10^{-3}$ | $1.8 \times 10^{-4}$ | HPLC-UV |
| pesticide | linuron | 330-55-2 | $C_9H_{10}Cl_2O_2N_2$ | 81 | 3.0 | $1.6 \times 10^{-4}$ | $5.1 \times 10^{-5}$ | HPLC-UV |
| pesticide | triazophos | 24017-47-8 | $C_{12}H_{16}O_3N_3PS$ | 30 | 3.3 | $4.1 \times 10^{-3}$ | $3.9 \times 10^{-4}$ | HPLC-UV |
| detergent | dodecylbenzene sulfonic acid sodium salt | 25155-30-0 | $C_{18}H_{29}NaO_3S$ | 800 | 0.5 | | | Spectro-photometer +MBAS |
| detergent | octaethylene glycol monotetradecyl ether | 27847-86-5 | $C_{30}H_{62}O_9$ | | ? | | | Potentiometric titration BiAS |
| organotin substances | tri-n-butyltin acetate | 56-36-0 | $C_{14}H_{30}O_2Sn$ | 65 | 3.2 | | | GC-ICPMS |
| organotin substances | triphenyltin acetate | 900-95-8 | $C_{20}H_{18}\ O_2Sn$ | | 3.4 | | | GC-ICPMS |
| various | humic acid sodium salt | 1415-93-6 | - | | | | | Not available |
| various | acetone | 67-64-1 | $CH_3COCH_3$ | | | | | GC-FID |

* Priority Substance, according to Water Frame Directive, Daughter directive, proposal 17 July 2006;
*** in solution, decomposes in UV light.

Henry's law constant can be calculated from vapour pressure by the formula H=(P*M)/(S*R*T) where H is Henry's law constant, P is vapour pressure, M is molecular weight, S is water solubility, R is the gas constant and T is temperature (Mensink et al., 2008).

# Appendix III. Calculation of the toxic pressure of metals

**Approach for determining toxic pressure of metals**
As already explained in chapter 3, metals do not adsorb onto XAD and are therefore not extracted. This implies that the toxicity of metals is not included in the experimentally determined toxicity. In the past, a few attempts were made to *calculate* the toxicity of metals en integrate that calculated value with the measured toxicity of the other substances. However, each of the applied metals has its disadvantages. Here we will go more deeply into this topic, and also we will discuss the most promising option for a future assessment of the toxic effects of metals.

Toxic pressure caused by metals in the aquatic environment is related to the biological availability, which is determined by the physico-chemical properties of the metal, the composition of the water and properties of the organisms in the ecosystem. Depending on the factors, a certain fraction of each metal is available to an organism. This biologically available fraction is variable. Environmental conditions may vary in time and space. For some factors the seasons play a role, but also during a day the composition of water may change due to e.g. photosynthesis. In addition, water is dynamic. Water flows from small brooks and ditches via rivers to the sea. Because the water composition changes during that transport, the biological availability of metals also changes (Griffioen and Van Weerdum, 2007, De Nijs et al., 2008).

**Facing the problem of estimating bioavailability**
Metal concentrations in surface waters are in general measured as so-called 'total' concentrations. The species in which metals occur, the so-called chemical speciation, is not measured on a routine-basis. Therefore it is not known which fraction of the metal concentration is bound to inorganic or organic complexes. However, the bioavailable fraction of the metal concentration, and hence the toxic pressure caused by a metal, is dependent on the chemical speciation. As stated before, for calculating the toxic pressure of metals and incorporating such a calculation into the pT-method, there is no approach available which is both scientifically sound and also widely accepted.

In the current methodology, as used e.g. by De Nijs et al. (2008), the toxic pressure is determined by the so-called ms-PAF method. That method uses toxicological data from laboratory tests to estimate the fraction of organisms that is likely to encounter adverse effects caused by elevated concentrations of metals in surface water. It is possible to correct the measured values in surface water for complexation by inorganic and organic complexes, which will lead to a lower availability of the metal. However, for the toxicological data from laboratory tests, the necessary data for such a correction generally is not available. To make it worse, the free concentration of metals in solution can not be translated directly to effects on organisms, caused by competition between metals or with other (macro) ions like calcium and magnesium.

Initially, the measured total concentrations in surface water were compared with toxicological data obtained from tests in synthetic media (Struijs et al., 2000). The disadvantage of this approach is that it does not take into account the chemical speciation of metals at all. Struijs and De Groot (2001) proposed to apply a new approach which does take into account chemical speciation of metals. In that new approach, the concentration of free metal in surface water was calculated, and then compared to toxicity data based on tests in synthetic media (e.g. De Groot et al., 2006). The disadvantage of this

approach is that it ignores the presence of inorganically complexed metal (like hydroxides and carbonates) in the synthetic media which *does* influence the toxicity data. Nevertheless, the approach of Struijs and De Groot (2001) was the best available at that time.

Since 2001, there was a lot of progress made in the knowledge of predicting behaviour of metals in the aquatic environment, especially the way chemical speciation should be dealt with. With current knowledge, it is no longer desirable to use approaches like the one proposed by Struijs and De Groot (2001). At present, better techniques are available.

**Possible solutions: Biotic Ligand Models**

In the past few years, so-called Biotic Ligand Models (BLMs) were developed to assess the effects of metals in the aquatic environment with a higher accuracy than was possible until recently. These BLMs are operational models that estimate the bioavailability and hence toxicity of metals in the aquatic environment. The models take into account the chemical speciation of metals in the water phase as well as binding of the metals to the organism, the so-called biotic ligand. Because of competition of a metal with other ions, the binding of a metal to an organism is complex. It depends on several factors, among which the composition of the surface water.

Figure III-1 shows the structure of a BLM. In the *chemical* part of the model it is estimated which part of the metal concentration ($M^{n+}$) is available for exchange with the organism (based on chemical equations and equilibria). In the *physiological* part it is estimated which part of the available metal concentration actually binds to the organism. In case of a fish it is the gill. Effects of a metal are related to the amount of metal actually bound to the organism.

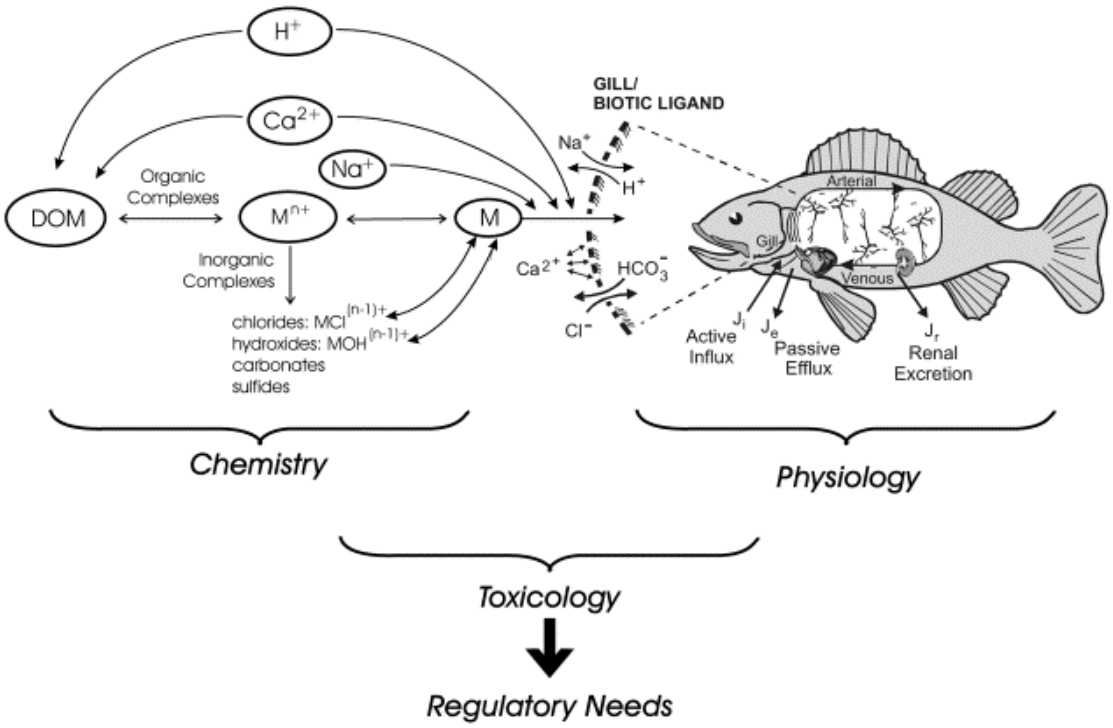


**Figure III-1. Schematic representation of the elements of a Biotic Ligand Model.**

**Available Biotic Ligand Models**
At present, BLMs are available and operational for copper, zinc and nickel. Preliminary results show that the fairly complex BLM-methodology for these three metals can be simplified for several types of surface waters. For lead, cadmium and cobalt, BLMs are under development.

BLMs allow an estimation of toxicological effects for each *individual* metal. At present, there is no method available for metal *mixtures*. Surface water is always a mixture of metals and as a consequence, the effects of metals in surface water can only be assessed on a per-metal basis. For the time being, the toxic pressure can be calculated per metal, using BLMs, and then added. For the moment, we assume that this approach gives a reasonable assessment of the actual toxic pressure of the mixture. Research is underway to focus on toxicity of *mixtures* of metals. Hopefully this will lead to a better method to assess toxicity of mixtures.

**Recommendations**
Methods that were applied previously to determine the toxic pressure of mixtures of metals in the pT-method, do not yield the required information in sufficient detail, but the developments with respect to BLMs are promising. It is recommended to integrate the developments in BLM-research into the pT-method. This implies that the following steps should be taken.

- Compile an inventory of existing validated BLMs, and make them operational within RIVM to estimate the toxic pressure of individual metals. This holds for copper, nickel and zinc.
- Pay attention to developments in new BLMs for cobalt, cadmium and lead.
- Investigate the results of existing and new models to come to methodologies to estimate toxicity of *mixtures* of metals. The calculated values should be integrated with the *measured* toxic pressure of organic pollutants.

With the knowledge we already have, it should be possible to develop new insights in the combined toxicity of copper, zinc and nickel in a few months time. Copper and zinc are very important causes of toxic pressure in surface water ((De Nijs et al., 2008). For other metals more time will be required.

**RIVM**

National Institute
for Public Health
and the Environment

P.O. Box 1
3720 BA Bilthoven
The Netherlands
www.rivm.com