Toxicity-Based Criteria for Receiving Waters Development of Sediment Toxicity Tests and Bioassays

R&D Technical Report
P315

Toxicity-Based Criteria for Receiving Waters
Development of Sediment Toxicity Tests and Bioassays

R&D Technical Report P315

R Fleming and I Sims

Research Contractor:
WRc plc

Further copies of this report are available from:
Environment Agency R&D Dissemination Centre, c/o
WRc, Frankland Road, Swindon, Wilts SN5 8YF

tel: 01793-865000 fax: 01793-514562 e-mail: publications@wrcplc.co.uk

Publishing Organisation:
Environment Agency
Rio House
Waterside Drive
Aztec West
Almondsbury
Bristol BS32 4UD

Tel: 01454 624400 Fax: 01454 624409

ISBN: 1 85705 1823

This report is the result of work jointly funded by the Scotland and Northern Ireland Forum for Environmental Research (SNIFFER) and the Environment Agency.

Dissemination status
Internal: Released to Regions
External: Released to the Public Domain

Statement of use
This project aimed to develop and standardise sediment toxicity tests and bioassays which would discriminate between sites of varying biological and chemical quality and give ecologically meaningful results. The report will be of interest to Environmental Protection Staff, Regional and Area Biologists and Ecologists, and is for information only.

Research contractor
This document was produced under R&D Project 560 by:

WRc plc
Henley Road
Medmenham
Marlow
Buckinghamshire
SL7 2HD

Tel: 01491 571531 Fax: 01491 579094

WRc Report No: EA 4299/1

Environment Agency Project Leader
The Environment Agency's Project Leader for R&D Project 560 was:
David Forrow/Oliver Warwick, National Centre for Ecotoxicology and Hazardous Substances

SNIFFER's Project Manager for R&D Project 560 was:
Dr J Redshaw, SEPA West

CONTENTS

APPENDICES

LIST OF TABLES

EXECUTIVE SUMMARY

The Environment Agency has put forward recommendations to use toxicity-based criteria to perform general quality classification of river and estuarine receiving waters and local impact assessment of specific inputs. For these purposes a battery of water and sediment toxicity bioassays is required which are shown to discriminate between sites of varying biological and chemical quality and give ecologically meaningful results. These bioassay methods must be easy to use and cost-effective when used on the scale required to perform national classification. In addition, the results must be repeatable in the same laboratory and reproducible between laboratories so that spatial and temporal bias is not introduced into the classification.

In this programme sediment bioassays were developed using sediment spiked with model toxicants in the laboratory and contaminated field sediments taken from pollution gradients. The primary test organisms were the larvae of the freshwater dipteran *Chironomus riparius* and the estuarine amphipod *Corophium volutator* for which acute tests, measuring lethality as a primary endpoint, have been developed and used successfully by the Environment Agency. Data generated with these species were compared with sediment toxicity test results using other test organisms including the marine polychaete worm, *Arenicola marina*, and the meiofaunal estuarine polychaetes, *Ophryotrocha* sp.

Chronic tests with *C. riparius* and *C. volutator* were more sensitive in terms of lethality than acute tests conducted previously with the same sediment and test substance. However, the chronic tests require additional expertise as they involve handling early life stages and measurement of sublethal endpoints. For this reason, reproducibility of these test results when tested in other European laboratories was not as pronounced as seen previously with acute tests. Both acute and chronic tests with these species should be tested on a wider range of field sediments before one can be selected in preference of the other. For this reason, both methods were recommended for inclusion in an Environment Agency pilot study performed in 1996.

When compared with other marine test species, *Corophium volutator* appeared to be the most sensitive to a selected model contaminant and more tolerant of difference in sediment type. In this species, lethality measured over a chronic test duration was more as sensitive as the sub-lethal endpoint of cast-formation measured in *Arenicola marina*. Cast-formation in *Arenicola marina* was more variable in coarse sediment types. However, because they could potentially supply supplementary information when used in conjunction with *Corophium volutator*, both methods were recommended for the Environment Agency pilot study.

The marine meiofaunal polychaetes investigated in this programme appeared not to be as sensitive as *Corophium volutator* to the field sediments tested. Further research is required on the relative sensitivity of these species to a range of chemicals and field sediments before they can be considered for regulatory purposes.

KEY WORDS

Sediment, toxicity, tests, bioassays, development

1. INTRODUCTION

The Environment Agency has put forward recommendations to use toxicity-based criteria to perform general quality classification of river and estuarine receiving waters and local impact assessment of specific inputs. For these purposes a battery of water and sediment toxicity bioassays is required which are shown to discriminate between sites of varying biological and chemical quality and give ecologically meaningful results. These bioassay methods must be easy to use and cost-effective when used on the scale required to perform national classification. In addition, the results must be repeatable in the same laboratory and reproducible between laboratories so that spatial and temporal bias is not introduced into the classification.

This report details a two year research programme which was designed to develop and standardise sediment toxicity tests and bioassays which would fulfil the above criteria. Short-term acute sediment tests have been developed for the Environment Agency in a previous programme. These tests have since been widely used for regulatory and research purposes and were submitted to the OECD and European Commission for consideration as draft standard guidelines. Acute sediment tests have been demonstrated to be a cost-effective, sensitive and reliable tool for assessing the toxicological effects of grossly contaminated sediments, such as may be encountered in local impact assessments. However, as mortality over a short-term exposure duration is the primary endpoint in such tests, they may not be sensitive enough to discriminate between the lower levels of sediment-bound contaminants likely to be present in receiving waters not directly impacted by a local discharge. Therefore, for general quality assessments at least, more sensitive and discriminatory measures are required which can detect toxicological effects at concentrations lower than those causing short-term lethality. This can potentially be achieved by extending the exposure duration of the acute tests, investigating sublethal as well as lethal endpoints, and exposing the most sensitive life-stages of the test organisms. Extending the exposure duration of sediment tests can potentially increase test sensitivity in three ways. First, the exposure of the animal is prolonged which increases the likelihood of chemical equilibrium being established between the sediment phase and the body tissue. In this way exposure of the laboratory test organism is maximised and more realistic compared with field organisms which are exposed to sediment-bound contaminants over extended durations. Secondly, all routes of contaminant exposure are taken into account. Acute exposure from sediment pore-water may be short-term, but exposure via ingestion and direct contact with sediment particles may take longer to mediate toxic effects. Thirdly, extending exposure duration increases the portion of the animals life cycle exposed and increases the probability of including sensitive life stages. Immature organisms often appear to be more sensitive to toxic agents than adult organisms, probably due to differences in the development of detoxification mechanisms (Rand *et al.* 1995) and so early life stages can be used to initiate tests. While these life stages may be difficult to recover from the sediment on termination of acute tests, after an extended duration they will usually be large enough to retrieve on sieving. Also, sensitivity may be increased at times of moulting or metamorphosis from larval to adult stages. The longer the test duration the more of these sensitive stages that will be exposed. An additional advantage is that if exposure occurs over a longer developmental time, the likelihood of measuring differences in growth and development in test animals is maximised.

The objectives of this programme were to develop long-term chronic sediment tests which could be used to measure both lethal and sublethal effects for classification of general receiving water quality and local impact assessment. Because the assessment of any test method involves inter-laboratory comparison to assess reproducibility, the programme was co-funded by the European Commission in order that laboratories in the Netherlands, Germany and Portugal could also become involved. WRc co-ordinated the programme.

The organisms used for test method development here were the two species for which acute tests had been developed previously: the freshwater dipteran *Chironomus riparius* and the estuarine amphipod *Corophium volutator*. These have since been recommended by the OECD for test guideline standardisation for the purposes of pesticide and general chemical risk assessment (OECD 1995). Additionally, the sensitivity of three other marine test organisms was investigated, including the marine polychaete worm, *Arenicola marina* as this is a species recommended by MAFF for surveys of marine and estuarine sediment quality and assessment of dredge spoil. In addition, a sub-contract was let to Royal Holloway University of London to assess the performance of the marine meiofaunal polychaetes *Ophryotrocha* sp. as sediment toxicity test organisms. This is included in Appendix A.

The specific objectives of the programme were as follows:

- to develop chronic (long-term) whole sediment tests for *Chironomus riparius* and *Corophium volutator*;
- to assess the reproducibility of both tests in inter-laboratory comparisons using sediment spiked with two model toxicants likely to be found in UK sediments;
- to assess the performance of these tests for consistently discriminating between field sediments taken from gradients of contamination;
- to compare results of chronic tests with acute tests developed previously;
- to compare the sensitivity and sediment tolerance of *Corophium volutator* with other marine species using different sediment types spiked with one model contaminant; and
- to develop a test method for exposure of *Ophryotrocha* spp. and to compare sensitivity of these species with other methods developed in the programme.

The other laboratories which participated in this study either funded by the European Commission or on a voluntary basis are as follows:

Rijkswaterstaat: Tidal Waters (RIKZ) (The Netherlands)

Rijkswaterstaat: Inland Waters (RIZA) (The Netherlands)

University of Utrecht (The Netherlands)

University of Hamburg (Germany)

Portuguese Institute for Investigation of the Sea (IPIMAR) (Portugal)

Ministry of Agriculture, Fisheries and Food (Directorate of Fisheries Research) (UK)

Zeneca Agrochemicals (UK)

University of Aberdeen (UK)

University of Aveiro (Portugal)

National Institute of Public Health and Environmental Protection, RIVM (The Netherlands)

TNO Institute of Environmental Sciences, Energy Research and Process Innovation (The Netherlands)

2. TEST METHOD DEVELOPMENT

2.1 Methods

2.1.1 *Chironomus riparius*

Whole sediment testing with *Chironomus* spp. was first reported by Wentsel *et al.* (1977). Due to its ease of culture (Nebeker *et al.* 1984b) the genus has been used previously in water column, porewater, elutriate and whole sediment testing ranging from 8 hour to 29 day exposures (Burton 1991), and an acute ten-day toxicity test method has been developed for the Environment Agency using *Chironomus riparius*. The larvae of the freshwater dipteran, *Chironomus riparius,* is widespread and abundant in European waters. It is a true sediment dweller and detritivore and thus is likely to be exposed to sediment-bound contaminants though all the major routes.

There are four distinct phases in the life cycle of *Chironomus riparius*: egg, larva, pupa, and adult midge. Under ambient conditions of 20 °C, good aeration and adequate food, this life cycle is usually completed within 30 days. Eggs are laid in gelatinous ropes (approximately 200 eggs per rope) just below the water surface and are usually attached to a solid object. Larvae hatch between 1-3 days after the eggs are laid, and larval growth and development occurs in four instar stages, each of which is about 4 to 7 days duration. The first instar is approximately 1 mm long and is a free-swimming pelagic life-stage; the second, third and fourth instars are approximately 1-3 mm, 3-6 mm and 6-20 mm respectively. Differences between developmental stage can be determined accurately by measuring head-capsule width which falls within discrete boundaries for each larval stage regardless of the condition of the organism. The fourth instar increases in weight as the larvae prepare to pupate. Because of these discrete developmental stages, sublethal measurements can be measured using both the larvae and the adults of this species.

The larvae of *Chironomus* spp. can tolerate very low levels of dissolved oxygen (Burton and MacPherson 1995) and thrive on many substrates ranging from filter paper to fine silt to coarse acid-washed sand. Therefore, they should tolerate a wide range of sediment types when exposed to a selection of field sediments. In addition, the species is sensitive to many contaminants, often as sensitive as *Daphnia* spp. for many organic and inorganic chemicals (Ingersoll and Nelson 1990). There is a database of toxicity data generated in both water and sediment exposures for different stages of larval development.

The larval stage, especially the first instar, is the most sensitive life-stage in terms of toxic effects (Burton *et al.* 1992), although the second and third instars are also widely used in toxicity testing. The second instar was used in acute tests developed previously. These are easier to handle as first instar larvae are very sensitive to air exposure, and may potentially become trapped in the surface tension of the water (Nebeker *et al.* 1984a, 1984b). As a pelagic life stage they also require additional feeding as they cannot utilise sediment organic carbon (Vine and Crane 1994). Eggs cannot be placed directly in sediment systems because abrasion

by sediment particles decreases hatching rate (Grootelaar 1995, pers. com.). Also, eggs are only available in gelatinous egg ropes which must be carefully dissected under a microscope to accurately count the numbers required for test inoculation.

With these considerations in mind two protocols for a chronic test with *Chironomus riparius* were proposed by RIZA in the Netherlands. Definitive protocols are shown in Appendix B, and summarised here:

Protocol 1: Egg ropes were exposed in sediment elutriate for seven days after which time 25 larvae were transferred to a sediment water system for a further 21 days or until emergence (if sooner). Mortality of larvae and differences between treatments in mean individual dry weight of fourth instar larvae were recorded. Differences in larval instar were also determined and the total number surviving to fourth instar and above was used as a sublethal endpoint in addition to dry weight.

Protocol 2: Egg ropes were dissected under a light microscope into batches of 25. These were then placed in plastic microcentrifuge tubes in sediment-water systems. After four days the tubes were checked for hatching success. If the eggs had not hatched they were replaced by a further batch of 25. Hatched larvae could swim out of the tubes directly into the sediment-water system, but any still in the tubes at day 4 were manually flushed into the system. The animals were exposed for 21 days or until emergence at which time larval mortality and differences in dry weight and larval instar were recorded.

In both Protocol 1 and Protocol 2 the animals were fed three times per week for the test duration with ground tetramin fish food solution: 100 μl of 0.5 mg tetramin in 25 ml deionised water per vessel for the first two weeks, 150 μl for the third week and 200 μl for the fourth week. The test systems were 100 ml tall-form Pyrex beakers with a sediment:water ratio of 1:4. Procedures for test termination were the same as in acute tests developed previously (Fleming *et al.* 1994) in that animals were sieved out of the sediment using a 250 μm sieve, killed in acetone and rinsed in deionised water. Prior to drying the numbers of survivors in each larval instar was assessed. This was done either by measuring individual head widths, which correspond to the four instars, or estimating instar by eye which is possible for experienced personnel. All fourth instar survivors in a treatment were then bulked, dried for 48 hours at 60 °C and weighed. Mean individual dry weights for fourth instar larvae were then calculated.

Test substance selection

The degree of binding to a sediment is dependent on the solubility of a test substance. For non-polar organic compounds the log Kow or Koc (octanol:water or organic carbon partition coefficient) can be used to determine whether the greater proportion of a substance will be bound to the sediment or present in the dissolved phase. The larger the Kow, the more likely the substance will be bound to sediment and it is generally accepted that a log Kow >3 indicates a high propensity to bind. For this study, lindane was chosen as a model toxicant because its log Kow value of 3.7 means that it will bind to sediment (Saleh *et al.* 1982) but will be easier to handle than other poorly-soluble substances. Also it is known to be stable in sediment-water systems for long periods of time (Caron *et al.* 1985). Lindane is the gamma

isomer of 1,2,3,4,5,6-hexachlorocyclohexane, a non-polar cyclic aliphatic compound widely used as an insecticide.

Permethrin is a commonly used synthetic pyrethroid insecticide with a solubility of 0.2 mg l^{-1} and a log Kow (octanol-water partition coefficient) of 6.6. These two properties mean that is strongly hydrophobic and will bind to the sediment to a greater extent than lindane. Permethrin bioaccumulates to high levels in aquatic organisms and has been shown to be highly toxic to fish, insects and algae. The mechanism of acute toxicity is inhibition of the enzyme acetyl-cholinesterase (AChE) responsible for the hydrolysis of the neurotransmitter acetylcholine in synaptic tissue. As a consequence, the neurotransmitter accumulates leading to hyperactivity and imbalances within the nervous system (Van Rijn *et al.* 1995). Permethrin has been found in sediments downstream of carpet-making factories and thought to be responsible for impoverished benthic fauna in these areas (Vine and Crane 1994).

Lindane inter-laboratory comparison

A natural 'uncontaminated' sediment taken from Newton Bay, Poole Harbour (OS SZ 001 845), was desalinated and spiked with lindane using the 'wet rolling technique' described previously (Fleming *et al.* 1994). This was ring tested according to the definitive protocols in Appendix B. Protocol 1 was tested in three laboratories using a common source of animals, and Protocol 2 was tested in two laboratories using in-house cultures at each. Sediment spiking was carried out at a single laboratory and all testing commenced at the same time to reduce variability in results caused by sediment ageing.

Permethrin inter-laboratory comparison

The same desalinated sediment was spiked with permethrin using an industrial food mixer (Crypto Peerless KNM6). The test compound was added to wet sediment (sieved to 500 μm) in an acetone carrier solution. Each treatment was mixed for one hour. This method allowed large volumes to be spiked at the same time and, therefore, reduced inter-replicate variability seen with the 'wet rolling' method used previously. Protocol 1 of the *C. riparius* long-term test was ring tested in five laboratories. The same animals were used in all laboratories and testing commenced on the same day to reduce variability in results due to sediment ageing. Protocol 2 was further refined and tested in one laboratory only using permethrin-spiked sediment. The same spiked sediments and animals were used for both Protocol 1 and Protocol 2 testing.

All statistical analysis was performed as specified in the protocol in Appendix B.

2.1.2 *Corophium volutator*

Corophium volutator is an abundant intertidal estuarine amphipod found along the north-east American and west European coasts (Ciarelli *et al.* 1997). It is a sediment dweller which preferentially ingests particulate material with high organic carbon content (Kukkonen and Landrum 1994) and is therefore likely to be exposed to sediment-bound contaminants though all major routes of exposure. It is easy to handle and is readily obtained either from the field or laboratory cultures (Lee 1977, Davies *et al.* 1991).

Amphipods have been identified as one of the most vulnerable species to contaminants in the estuarine benthic ecosystem being one of the first group of organisms to disappear from benthic communities in sediments impacted by pollution (Swartz *et al.* 1982). Many species, including *Corophium volutator*, have been shown to be sensitive to contaminants such as hydrocarbons (Matthiessen and Thain 1989), heavy metals (Erdem and Meadows 1980, Bryant *et al.* 1985), and PAHs (Plesha *et al.* 1988, Reichart *et al.* 1985). *Corophium volutator* is also tolerant of wide ranges of abiotic factors such as temperature, salinity (Mills and Fish 1980, Bryant *et al.* 1984), substrate type (Sims and Fleming 1994, Davies *et al.* 1991, Roddie *et al.* 1991, Meadows 1964, Meadows and Reid 1964) and sediment sulphide (Meadows *et al.* 1981).

C. volutator has a longer life cycle than *C. riparius*. It has two generations per year, the first hatching from March to June reaching maturity two months later, and the second hatching in June to October and overwintering to continue to reproductive cycle in the following Spring (Ciarelli *et al.* 1997). It is therefore not cost-effective or practical to expose these animals from their early life-stages to adult in order that a large proportion of their life cycle can be investigated. An additional difficulty with this organism is that is does not have discrete developmental stages as does *C. riparius*, and so appropriate sublethal endpoints are more difficult to define. In a previous study, in which differences in body length over a 28 day period were measured, no significant differences were seen on exposure to a concentration range of lindane (Fleming *et al.* 1993). Therefore, body length is not considered as a suitable endpoint although is has been used as a measure of sublethal toxicity for other amphipod species (Dewitt *et al.* 1992). The growth endpoint, as measured by differences in mean individual dry weight, has been the only sublethal endpoint identified for this species (Ciarelli, RIKZ, the Netherlands, pers. com. 1995). In order to attain the maximum growth rate over the test duration, juvenile animals can be exposed during a fast developmental phase.

A 28 day test was proposed by RIKZ in the Netherlands using juvenile animals between 500 and 600 μm long, as measured by movement through sieves of this size mesh. The test systems (in terms of vessel size and sediment volume) were the same as those used in acute tests developed previously (Fleming *et al.* 1994). Throughout the exposure period no additional feeding was provided. Survival was assessed, and growth, as measured by mean individual dry weight, was the sublethal endpoint. The definitive protocol is given in Appendix B.

Lindane inter-laboratory comparison

Two natural 'uncontaminated' sediments were spiked with lindane using the 'wet-rolling' technique developed previously (Fleming *et al.* 1994). Spiking of each sediment was carried out by a single laboratory. Sediment 1 was that used previously for development of acute toxicity tests (Fleming *et al.* 1994) and for the *C. riparius* testing above. Sediment 2 had an organic carbon content approximately four times less than Sediment 1 and was taken from an uncontaminated site in the Netherlands. These spiked-sediments were ring-tested according to the definitive protocol in three laboratories. The same source of juvenile animals was used in all laboratories and all testing commenced at the same time to reduce variability in results caused by sediment ageing.

Permethrin inter-laboratory comparison

A single natural 'uncontaminated' sediment (Sediment 1) was spiked with permethrin using an industrial food mixer as described above. This sediment was ring-tested according to the definitive chronic protocol in five laboratories. The same source of juvenile animals was used in four laboratories whilst the fifth laboratory used an in-house source. All testing commenced at the same time to reduce variability in results caused by sediment ageing.

All statistical analysis was performed as specified in the protocol in Appendix B.

2.2 Results

2.2.1 *Chironomus riparius*

Lindane inter-laboratory comparison

Protocol 1

Effects concentrations for Protocol 1 using lindane-spiked sediment (LC_{50} and NOEC values), can be seen for the laboratories in Table 2.1. All concentrations were measured per gramme dry weight sediment.

Table 2.1 LC_{50} and NOEC values for Sediment 1 spiked with lindane - *C. riparius*, Protocol 1

Effect concentration ($\mu g\ g^{-1}$)	Lab 1	Lab 2	Lab 3	Lab 4
LC_{50}	0.27 (0.22-0.33)	0.23 (0.21-0.25)	**	1.03 (0.87-1.21)
NOEC (mortality)	<0.14	0.14	0.14	0.29
NOEC (dry weight)	*	*	<0.14	-
NOEC (development)	<0.14	0.14	0.14	-

* no significant difference between treatments and control (p=0.05)
** invalid

Dose response relationships were seen for survival in all four laboratories, but in Laboratory 3 the control survival was low, making the test invalid. Survival was higher in all treatments in Laboratory 4. This was because the sediment was diluted prior to testing which would have led to reduced test concentrations.

Dry weight and development data were not reported for Laboratory 4. In Laboratories 1 and 2, mean individual dry weights increased with increasing lindane concentration. This could be due to decreased competition for food in these vessels with increasing mortality. A dose response relationship was seen for the development endpoint, with the number of larvae at fourth instar and above decreasing with increasing lindane concentration. Development in Laboratory 3 was lower than in Laboratories 1 and 2 but this probably reflected the higher mortality seen at this laboratory. In all cases, development was a more sensitive endpoint than survival and more reliable than dry weight as it was not influenced by the number of survivors.

Protocol 2

In both laboratories which used Protocol 2, the control survival was low therefore making the tests invalid. Also, variability was high between replicates. Dry weight increased with increasing lindane concentration in both tests.

Permethrin inter-laboratory comparison

Protocol 1

Effects concentrations for Protocol 1 using permethrin-spiked sediment (LC_{50} and NOEC values) can be seen in Table 2.2 for the five participating laboratories. All concentrations were measured per gramme dry weight sediment.

Table 2.2 LC_{50} and NOEC values for Sediment 1 spiked with permethrin - *C. riparius*, Protocol 1

Effect concentration ($\mu g\ g^{-1}$)	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
Duration (days)	28	28	28	25	26
LC_{50} (95% CLs)	2.51 (2.28 2.76)	4.05 (3.58-4.58)	1.74 (1.56-1.93)	*	1.66 (1.45-1.90)
NOEC (mortality)	1.09	2.00	0.51	2.00	0.51
NOEC (dry weight)	**	**	**	**	**
NOEC (development)	2.00	2.00	0.51	2.00	0.51

* invalid
** increase in dry weight seen compared to control

Laboratory 4 had low control survival, therefore making this test invalid. In all other laboratories control survival was acceptable and dose response relationships were seen.

In all laboratories, mean individual dry weight increased with increasing permethrin concentration which was again related to an increase in mortality. Laboratory 1 had a higher control weight then the other laboratories. This is probably because 20 animals were used rather than 25, reducing competition for food.

Dose response relationships were seen for development in all laboratories

Protocol 2

Unlike the Protocol 2 tests with lindane-spiked sediment, the control survival was valid with permethrin-spiked sediment, and the within-replicate variability was reduced in the one laboratory that carried out this test. The 28 day LC_{50} was 3.82 (3.46-4.23) $\mu g\, g^{-1}$ permethrin. When this is compared to the LC_{50} value of 4.05 $\mu g\, g^{-1}$ for Protocol 1 achieved in the same laboratory (Laboratory 2), it can be seen that Protocol 2 was slightly more sensitive in terms of mortality although this is not significant as can be seen from the 95% confidence limits.

2.2.2 *Corophium volutator*

Lindane inter-laboratory comparison

Sediment 1

For Sediment 1 spiked with lindane, effects concentrations can be seen in Table 2.3 for the three participating laboratories. All concentrations were measured per gramme dry weight sediment. Dose response relationships were seen for survival in all laboratories. In Laboratory 3 large variability was seen between replicates and control mortality was high. This may have been due to transportation stress or to specific conditions within the laboratory. High control mortality and large variability had been seen previously in this laboratory during inter-laboratory comparisons of short-term tests (Fleming *et al.* 1994). Differences in survival were also seen between Laboratories 1 and 2 which is shown by the disparity in LC_{50} values.

Table 2.3 LC_{50} and NOEC values for Sediment 1 spiked with lindane - *C. volutator*

Effect concentration ($\mu g\, g^{-1}$)	Lab 1	Lab 2	Lab 3
28 day LC_{50} (95% CLs)	2.49 (1.84-3.37)	0.97 (0.69-1.34)	0.26 (0.19-0.35)
NOEC (mortality)	1.52	0.18	*
NOEC (mean dry weight)	*	*	*

* no significant differences between treatments and control (p=0.05)

A dose response trend was seen for mean individual dry weight in all laboratories, with a decrease in dry weight with increasing lindane concentration, although in the highest concentration an increase in dry weight was observed in two laboratories as seen in the *C. riparius* tests. However, differences in dry weights determined at all laboratories were not statistically significantly different from the mean control values.

Sediment 2

For Sediment 2 spiked with lindane, effects concentrations for the three participating laboratories can be seen in Table 2.4. All concentrations were measured per gramme dry weight sediment. All tests were valid in terms of control survival, but LC_{50} values were outside the concentration range tested in Laboratories 1 and 2. The measured concentration in the highest treatment was 0.75 $\mu g\ g^{-1}$.

Table 2.4 LC_{50} and NOEC values for Sediment 2 spiked with lindane - *C. volutator*

Effect concentration ($\mu g\ g^{-1}$)	Lab 1	Lab 2	Lab 3
28 day LC_{50} (95% CLs)	**	**	0.68 (0.50-0.93)
NOEC (mortality)	*	0.24	*
NOEC (mean dry weight)	0.41	0.04	0.49

* no significant differences between treatments and control (p=0.05)
** LC_{50} outside concentration range tested

In all three laboratories a significant decrease (p=0.05) was seen in mean individual dry weight with increasing concentration. In Laboratory 2 there was an increase in dry weight at the highest concentration tested, as seen in the Sediment 1 tests.

Permethrin inter-laboratory comparison

Effects concentrations for permethrin-spiked sediment can be seen for the four participating laboratories in Table 2.5. These were based on nominal concentrations as analytical confirmation showed that two of the concentrations tested were not significantly different from nominals. Analysis of the other test concentrations did not give interpretable results due to poor recovery during analysis and impurities in the sample.

Table 2.5 LC_{50} and NOEC values for Sediment 1 spiked with permethrin - *C. volutator*

Effect concentration ($ng\ g^{-1}$)	Lab 1	Lab 2	Lab 3	Lab 4**
28 day LC_{50} (95% CLs)	71.52 (66.72- 77.24)	58.70 (53.06-64.99)	48.77 (42.10-56.51)	67.00 (55.00-82.00)
NOEC (dry weight)	80	*	*	-

* no significant differences between treatments and control (p=0.05)
** different source of animals

The LC_{50} values for this test were less variable than for the lindane test. The reproducibility of results can be assessed by considering the Coefficient of Variation of the LC_{50} values generated. The Coefficient of Variation (CV) is the ratio of the sample standard deviation to the mean. High CVs for LC_{50} values reflect high test variability and poor precision. A CV of below 30% is generally felt to be acceptable for a standard protocol (Whitehouse and Delaney 1995). Environment Canada (1990) have specified a CV of 20% to be acceptable although they state that a CV of 30 may be more realistic for some tests. The CV for the LC_{50} values here was 16.28% which indicates a relatively high degree of reproducibility. This is the case, even though two different sources of test animal are used.

In Laboratories 3 and 4, higher dry weights were seen at top concentrations when compared to the control. This is the same result seen in the lindane test. In Laboratory 1, however, a significant decrease in weight was seen at 80 $ng\ g^{-1}$ although this was above the LC_{50} concentration. In Laboratory 2, a decrease in dry weight was also seen at this treatment compared to controls although this was not significant at p=0.05.

2.3 Discussion

2.3.1 *Chironomus riparius*

Reproducibility

For Protocol 1 with lindane-spiked sediment, variability in survival was seen between the four laboratories. This can partly be explained by the fact that Laboratory 4 diluted the sample prior to testing. This demonstrates the importance of including explicit details on sediment-preparation and sample handling in the test protocol. Also this could have been due to differences in the actual lindane concentrations tested. Using the 'wet rolling' method, replicates were spiked in individual bottles. This may have led to variability between replicates and is particularly a problem when effects concentrations are based on the measured

concentration of a composite sample rather than the measured concentration in each individual sample. This was rectified in later testing by using a food mixing method in which larger volumes of sediment could be spiked in batches.

Of all the endpoints measured, development, (i.e. number of individuals at fourth instar and above) was the most sensitive used. It was more reliable than mean individual dry weight which was influenced by the number of survivors. Where significant differences in dry weight were seen relative to controls, NOEC values were higher than LC_{50} values. Therefore the dry weight endpoint was not suitable for measuring sublethal effects for this test substance.

For the permethrin test using Protocol 1, there was some variability between laboratories in terms of mortality. The LC_{50} value at Laboratory 2 suggested lower toxicity than determined at Laboratories 1, 3 and 5. A possible explanation for this is that the animals were supplied from Laboratory 2, and sub-cultures transported to the other laboratories may have suffered some transportation stress. Dry weight increased in all laboratories with increasing concentration showing once again, that this may be an unsuitable sublethal endpoint. As in the lindane tests development was once again a more sensitive and reliable endpoint. Therefore development is recommended as the primary sublethal endpoint in further testing with this test protocol.

For the lindane test using Protocol 2, high control mortality and variability between replicates invalidated the results. Protocol 2 was refined following the lindane test and results showed that control survival in the permethrin study was valid and variability between replicates reduced. The Protocol 2 test was apparently more sensitive than Protocol 1 for permethrin - spiked sediment, but not significantly so. This test was not carried out in other laboratories with less experience. However, given that the results were not significantly more sensitive, and that the protocol requires more skill than Protocol 1, it is recommended that Protocol 1 should be used as the preferred protocol of the two unless more development and evaluation is carried out with Protocol 2.

Refinement is required to reduce inter-laboratory variability of these test results and to improve control survival in some laboratories. Some degree of experience is required to conduct these tests, and the variability reported here (other than that caused by differences in the test media) may be a reflection of this. In general, reproducibility was best (particularly in the permethrin test) in Laboratories 1 and 2 which had the most experience with this test organism.

Acute versus chronic tests

Table 2.6 shows the lindane effects concentrations measured in the chronic test (Protocol 1) and the acute ten day test carried out previously with lindane (Fleming *et al.* 1994). Both tests were performed in two laboratories using the same sediment.

Table 2.6 Comparison of effects (lindane $\mu g\ g^{-1}$) for acute and chronic *C. riparius* tests

	10 day LC_{50} (acute test) (95% CLs)	28 day LC_{50} (chronic test) (95% CLs)	NOEC (mort) (chronic test)	NOEC (dw) (chronic test)	NOEC (dev) (chronic test)
Lab 1	0.72 (0.61-0.84)	0.27 (0.22-0.33)	<0.14	*	<0.14
Lab 2	0.60 (0.54-0.67)	0.23 (0.21-0.25)	0.14	*	0.14

* No significant decrease in dry weight compared to controls

The chronic test used here (Protocol 1) was more sensitive than the previous acute test in terms of mortality by a factor of 2.6 in both laboratories. The development endpoint in the chronic test also showed effects at lower concentrations than the LC_{50}. However, the NOEC for both mortality and development was the same. This may, however, reflect the difficulties in deriving an NOEC value which is determined by the test concentration used.

The differences in sensitivity between acute and chronic tests for lindane were smaller than would be expected given the extended duration, early life stage and sublethal endpoints. This may be due to the toxic mode of action of lindane. The relative sensitivity of the two tests should be demonstrated for a wider range of chemicals before chronic tests with this species can be used in place of acute tests. The chronic tests require more experience of the test animal, and require more resources and to carry out. If the difference in sensitivity is not large, acute tests may be more cost-effective. In addition, one of the problems associated with the chronic test method developed here is that the eggs must undergo a period of pre-exposure in elutriate after which there is no quantifiable endpoint. This pre-exposure stage may bias the results of the sediment test as the survivors of the elutriate exposure may be less sensitive than the rest of the test population. This protocol is, however, more realistic of field scenarios, when eggs and first instar larvae are exposed in water before moving into the sediment, and while it may not be ideal for toxicity testing in risk assessments, it may be more suitable for bioassays of field sediments (Section 3). These points should be considered during selection of methods for Environment Agency purposes.

2.3.2 *Corophium volutator*

Reproducibility

Variability in survival of *C. volutator* between laboratories was high for both sediments spiked with lindane using the 'wet rolling method'. As with the *C. riparius* test this could be due to the spiking method used rather than the toxicity test methods. For the permethrin test, in which the food mixer was used for spiking, variability in the survival endpoint of the

C. volutator test was reduced, giving a CV of 16.28% for four laboratories, one of which had used a different source of animals. This highlights the need for a homogeneously spiked sediment when conducting inter-laboratory testing, and the need for confirmatory chemical analysis in risk assessments employing these tests.

Sediment type

For the lindane test, mean individual dry weight was higher after a 28 day exposure in Sediment 2 than in Sediment 1. This was unexpected given that the organic carbon content (the expected food source) of Sediment 2 was four times less than Sediment 1. Comparison of mean individual weights before and after exposure in Sediment 1 showed that the animals had not grown in this sediment, and inactivity was seen throughout the test period.

Given that Sediment 1 did not support growth over the test duration it is possible that this sediment contained low levels of toxicants or other stressors. However, results of previous tests with the same sediment in which the animals were given additional food (Fleming *et al.* 1994), show that growth rates in this sediment were the same as Sediment 2. Therefore, it is likely to be a case of food availability. It is possible that the organic carbon type of Sediment 1 was not available to the organisms as a food source.

In Sediment 2, significant differences were seen in individual dry weight at concentrations lower than those causing mortality in all three laboratories. This was not seen in Sediment 1. Therefore, dry weight may be a sensitive sublethal endpoint to include in chronic testing only when the test sediment used promotes growth. This may not be a problem for risk assessment where a standard and optimised sediment could be used in all tests. However, for assays of field sediment, where each sediment will contain different levels and types of organic carbon content, growth may not be a suitable endpoint. The other problem in using individual dry weight as an endpoint is that mean values tend to increase as mortality increases, possibly due to decreased competition for food. Therefore, it may only be useful in these tests systems for those sediments which are not acutely toxic enough to cause lethal effects. An alternative would be to expose animals singly. If dry weight is to be used, a validity criterion, specifying acceptable levels of mortality, should be included in the protocol.

A potential way to overcome problems with food availability is to feed the animals with additional carbon throughout the test duration. Although additional feeding could increase or decrease contaminant bioavailability, depending on the contaminant, the test may be more discriminatory between test concentrations for the dry weight endpoint. This should be considered in further test method development.

Survival data for Sediments 1 and 2 show that in terms of lethality, the toxicity of lindane in both was similar, even though their organic carbon contents were very different. This is an unexpected result given that organic carbon is thought to be one of the major factors controlling bioavailability (DiToro *et al.* 1991), and toxicity would be expected to be higher in the sediment containing less organic carbon. These results show that organic carbon type could be as important as organic carbon content in determining toxicity. The factors controlling bioavailability are complex and are the focus of another research programme underway at WRc funded by MAFF (Boxall and Fleming 1997).

Acute versus chronic tests

Table 2.7 shows the lindane effects concentrations for the chronic protocols reported here, and the acute protocols (ten day exposure period) performed in the previous programme (Fleming *et al.* 1994). Both tests were carried out using the same sediment (Sediment 1) and the same source of test animals.

Table 2.7 Comparison of effects concentrations (lindane $\mu g\ g^{-1}$) for acute and chronic *C. volutator* tests

	10 day LC_{50} (acute test) (95% CLs)	28 day LC_{50} (chronic test) (95% CLs)	28 day NOEC (mortality) (95% CLs)	28 day NOEC (dry weight) (95% CLs)
Lab 1	4.18 (2.89-6.03)	2.49 (1.84-3.37)	1.52	*
Lab 2	2.97 (2.39-3.70)	0.97 (0.69-1.34)	0.18	*
Lab 3	2.25 (1.79-2.83)	0.26 (0.19-0.35)	*	0.29

* no significant differences between treatments and control (p=0.05)

These results show that chronic *C. volutator* tests are more sensitive than acute tests in terms of mortality although the difference between acute and chronic results generated in separate laboratories are different. As for the *C. riparius* tests, the relative sensitivity of the two tests should be demonstrated for a wider range of chemicals before chronic tests are recommended in place of acute tests. As discussed above, dry weight may not be a suitable sublethal endpoint for this species as it is strongly influenced by food availability which is probably linked to organic carbon type as well as content. Standardised sediment types for risk assessments should be selected with optimum growth in mind if this endpoint is to be used. Other endpoints such as reproduction could also be investigated further.

3. FIELD SEDIMENTS

3.1 General

One of the major differences in testing field-contaminated sediments, as opposed to sediments spiked in the laboratory with single chemicals, is that every field sediment differs not only in number and quantity of contaminants, but also in physico-chemical character such as particle size distribution and organic carbon content and type. These sediment characteristics may render the toxicity tests developed in Section 2, unsuitable for use as bioassays of field sediments.

The methods described in Section 2 were developed primarily for the testing of laboratory spiked-sediments. The objective of the work described in the section was to determine whether the same methods could be used to consistently discriminate between, and rank field-contaminated sediments in order of toxic effect, in different laboratories. This would give some idea of the validity of using these protocols for field assessments of sediment toxicity.

3.2 Methods

3.2.1 *Chironomus riparius*

Six freshwater field sediments, including a control, were collected from sites in the Netherlands ranging from uncontaminated to grossly contaminated (Grootelaar pers. comm. 1996). Protocol 1, as developed in Section 2 (see Appendix B) was used to pilot test the sediments in Laboratory 2 before they were used in an inter-laboratory comparison by five laboratories. In Laboratory 1, an acute ten day test was also carried out on the same sediments according to the protocols developed previously (Fleming *et al.* 1994).

3.2.2 *Corophium volutator*

Five estuarine field sediments were collected from sites in the Netherlands which had previously been classified as ranging from uncontaminated to grossly contaminated (Ciarelli pers. comm. 1996). Sediments were collected by grab sampler and the top 50-80 cm were used for testing. The samples were homogenised, but not sieved, and stored in the laboratory at 4 °C. Ten day tests were carried out with *C. volutator* at Laboratory 2 to determine acute toxicity and to demonstrate an 'acute toxicity ranking'. The sediments were then ring-tested in four laboratories according to the definitive chronic protocol (see Appendix B). All tests commenced on the same day to reduce variability caused by sediment ageing. The same natural 'uncontaminated' sediment, as used in Section 2, was used as a control sediment.

3.3 Results

3.3.1 *Chironomus riparius*

Pilot test

Mean survival, individual dry weight and development data can be seen in Table 3.1 for the pilot test using chronic Protocol 1 in Laboratory 2.

Table 3.1 Mean survival, dry weight and development data for pilot tests of *C. riparius*, Protocol 1 at Laboratory 2 using field sediments

Sediment	Mean % survival	Mean individual dry weight (mg)	Development %
Control	94	0.57	93
MAR	92	0.62	92
KET	88	0.63	86
BRB1	72	0.76	70
BRB2	95	0.54	93
DOM	74	0.63	74
HRV	89	0.46	86

Survival was significantly different (p=0.05) in sediments BRB1 and DOM. The same pattern was demonstrated by development, but for mean individual dry weight there were no significant differences between sediments as would be expected from the results of spiked-sediment tests in Section 2.

Inter-laboratory comparison

Two sediments, BRB1 and DOM, along with the control sediment, were chosen for the inter-laboratory comparison exercise. Duplicates of each were tested in five laboratories. Mean survival, dry weight and development data can be seen in Tables 3.2, 3.3 and 3.4 respectively.

Table 3.2 Survival data for chronic tests with field sediments - *C. riparius*

Sediment	Mean % survival				
	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
Control	84	98	87	90	91
DOM (1)	84	83	92	77	84
DOM (2)	88	86	85	65	88
BRB1 (1)	72	95	89	79	90
BRB1 (2)	79	96	84	81	93

Table 3.3 Dry weight data for chronic tests with field sediments - *C. riparius*

Sediment	Mean individual dry weight (mg)				
	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
Control	0.62	0.55	0.74	0.56	0.40
DOM (1)	0.67	0.73	0.58	0.50	0.47
DOM (2)	0.63	0.69	0.61	0.68	0.43
BRB1 (1)	0.65	0.66	0.66	0.79	0.36
BRB1 (2)	0.68	0.66	0.61	0.67	0.36

Table 3.4 Development data for chronic tests with field sediments - *C. riparius*

Sediment	Mean % development				
	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
Control	55	96	80	16	60
DOM (1)	71	83	87	43	83
DOM (2)	72	84	81	27	84
BRB1 (1)	55	91	81	27	78
BRB1 (2)	56	88	76	31	88

There were no consistent trends and no significant differences between sediments for any of the endpoints tested. This is possibly because the sediments chosen for the test did not cover a wide enough range of toxicity and the chronic test was not able to discriminate between them. The differences in toxicity, as measured by both lethal and sublethal endpoints in the pilot test, were not observed in the inter-laboratory comparison. This may be due to the fact that the sediment were stored between testing times. This highlights the need for immediate testing of field sediments following sampling. Ideally, this comparison should be repeated with a wider range of sediments.

Coefficients of variation (CV) between laboratories have been calculated for each sediment, for each of the endpoints tested. These can be seen in Table 3.5.

Table 3.5 Coefficients of variation between laboratories for the three endpoints in the Protocol 1, *C. riparius* test

Sediment	CV for mortality	CV for dry weight	CV for development
Control	6	22	49
DOM (1)	6	19	25
DOM (2)	12	17	35
BRB1 (1)	11	26	39
BRB1 (2)	9	23	36

This shows that survival is the least variable endpoint, and development the most variable endpoint. CVs for survival are very low compared to some of the results found previously in the spiked sediment toxicity tests, due to the low toxicity observed for each of the three sediments tested here.

Mean survival and dry weight data can be seen in Table 3.6 for the acute ten day test carried out at Laboratory 1 with the same sediments.

Table 3.6 Survival and dry weight data for acute tests carried out with field sediments

Sediment	Mean % survival	Mean individual dry weight (mg)
Control	95	0.72
DOM (1)	98	0.79
DOM (2)	93	0.62
BRB1 (1)	98	0.51
BRB1 (2)	95	0.60

Again, no significant differences or consistent trends were seen between the sediments.

3.3.2 *Corophium volutator*

Acute pilot test

Mean survival and dry weight data for the acute pilot test carried out at Laboratory 2 can be seen in Table 3.7.

Table 3.7 Mean survival and dry weight data for acute pilot test at Laboratory 2 - *C. volutator*

Sediment	Mean % survival	Mean individual dry weight (mg)
Control	94	0.55
M12	97	0.24
M13	10	0.14
M28	97	0.20
M7	100	0.33
M6	18	0.25

These sediments showed a good range of toxicity for both the lethal and the sublethal endpoint. Dry weight of animals in all of the field sediments were significantly lower than control weights (p=0.05), even for those sediments in which mortality was not significantly different from controls.

Chronic inter-laboratory comparison

Mean survival and dry weight data for the chronic test can be seen for the four participating laboratories in Tables 3.8 and 3.9.

In Laboratory 3, the control mortality was below the maximum acceptable limit for the test, thus making the these results invalid. This was the same laboratory in which high control mortality was also seen in spiked-sediment toxicity tests described in Section 2. All laboratories found a significant decrease in survival for M13 compared with the control. For Laboratories 1 and 2, significant differences were also found in M6. (Laboratory 3 also found significant differences from the control in M12 and M8, but the validity is questionable because of poor control survival). Toxic effects in M13 and M6 are consistent with the results of the acute pilot study although toxicity in M6 may have been reduced in the chronic test, possibly due to storage.

Table 3.8 Mean survival for chronic tests with field sediment - *C. volutator*

Sediment	Mean % survival			
	Lab 1	Lab 2	Lab 3	Lab 4
Control	82	94	77	96
M12	75	85	30	93
M13	0	8	18	25
M28	83	87	47	93
M7	70	95	62	95
M6	38	27	55	90

Table 3.9 Dry weight data for chronic tests with field sediments - *C. volutator*

Sediment	Mean individual dry weight			
	Lab 1	Lab 2	Lab 3	Lab 4
Control	0.37	0.48	-	0.66
M12	0.32	0.58	-	0.81
M13	-	0.49	-	4.41
M28	0.27	0.27	-	0.08
M7	0.26	0.32	-	0.82
M6	0.29	0.33	-	0.62

The dry weight results from Laboratory 3 were an order of magnitude higher than the other three laboratories, suggesting that an incorrect food ration was used. These data are therefore invalid and have not been included. Results for the remaining laboratories were variable, with the exception of sediment M28 where a decrease in growth was seen in all cases. This decrease was statistically significant in Laboratories 2 and 4, and is not consistent with the mortality data which showed no differences between Sediment M28 and the control in any laboratory.

Laboratory 1 found a decrease in growth compared to the control for all sediments except M12. A comparable trend was seen in Laboratory 2 although the differences were not significant at p=0.05. Laboratory 4 showed significant differences to these two laboratories for three of the sediments.

Coefficients of variation for each endpoint in each sediment are shown in Table 3.10. The highest CVs were found for those sediments which caused the greatest mortality. The results for Laboratories 1 and 2 were very reproducible as no significant differences were found between them for any of the sediments, for all endpoints measured. In these two laboratories the toxicity ranking of the sediments was the same and corresponded to the ranking shown in the acute pilot study.

Table 3.10 Coefficients of variation between laboratories for *C. volutator*

	CV for mortality	CV for dry weight
Control	10	30
M12	40	42
M13	86	-
M28	27	52
M7	21	66
M6	52	44

3.4 Discussion

3.4.1 *Chironomus riparius*

Unfortunately the differences in toxicity between the three sediments tested were insufficient for the chronic *C. riparius* test to discriminate between them. Differences were observed in both lethal and sublethal endpoints in the pilot test but it is possible that toxicity may have decreased during subsequent storage. However, results in all laboratories were similar and coefficients of variation, particularly for survival, were small compared to the spiked sediment toxicity tests. This was due to the low levels of toxicity seen. The sensitivity of the chronic protocols compared with the acute protocol developed previously (Fleming *et al.* 1994) cannot be ascertained on the basis of these data. This should be investigated in further programmes. However, no further discrimination was seen using the chronic protocol than using the acute method.

3.4.2 *Corophium volutator*

Acute versus chronic tests

Results from this test showed that the field sediments had a suitable range of toxicity for the inter-laboratory comparison of the chronic protocol. Sediments M13 and M6 were the most toxic in terms of lethality.

In short-term tests, growth (as measured by dry weight) was a more sensitive endpoint than mortality as significant differences in dry weight were seen when compared with controls, even when there was no significant mortality.

A comparison of the acute and chronic test conducted at Laboratory 2 show that the two gave similar results with the exception of sediment M6. In sediment M6, the mortality was not as pronounced in the chronic test as it had been in the acute test. This could have been due to a loss of toxicity over the storage time as there was several weeks delay between the pilot test and the inter-laboratory comparison. This has been seen in previous studies where loss of toxicity in metal contaminated sediments has been seen, although this is not as pronounced for organic chemicals (Burton and MacPherson 1995). A loss in toxicity was not seen for sediment M13.

In terms of dry weight, a comparison of acute and chronic tests showed that two of the five sediments tested were not significantly different. For the remaining three sediments, the dry weight endpoint in the chronic test was less discriminatory than in the acute test. Over the chronic test duration, increases in dry weight compared to the control were seen, whereas only decreases were seen over the acute test duration. These results suggest that dry weight was a more sensitive endpoint over the acute exposure duration, where significant differences were seen in all sediments compared with the control. One possible explanation is that over the longer term, food availability became a limiting factor and the content or quality of carbon in each field sediment influences growth to a greater extent than did contaminant load. As the only advantage to a chronic exposure duration is the increased mortality, it may be that acute tests are more cost-effective. However, more data are required before this can be concluded, and the affect of supplying additional food to enhance discrimination of the dry weight endpoint should be investigated further. Further comparisons between acute and chronic protocols should be carried out for field sediments, focusing on the need for additional feeding.

Reproducibility

For the chronic tests, mortality was the most reproducible endpoint with all laboratories identifying the same sediment (M13) as significantly more toxic in terms of lethality than the control. The dry weight endpoint was more variable although a general decrease in growth was seen in all laboratories in sediment M28. For both endpoints, good reproducibility was seen in two of the laboratories, with no significant differences between them for any of the sediments tested. In these two laboratories, sediment M6 was also identified as significantly different from controls in terms of lethality, and a decrease in dry weight was seen in four sediments (although this was only significant in one of the two laboratories). These two laboratories were those with the most experience of handling and testing sediments and *Corophium*, thus highlighting the need for experienced personnel in order to obtain reproducible results.

4. COMPARISON OF FOUR MACROINVERTEBRATE SPECIES

4.1 Objectives

As discussed previously, two requirements of sediment toxicity tests for national classification as that they are sensitive/discriminatory and are tolerant of a wide variety of sediment types. The objective of this study was to select potential sediment toxicity test species from the existing literature and to compare their performance in terms of control survival in different sediment types, and sensitivity to lindane.

4.2 Methods

4.2.1 Test species

The following species were selected for this study:

Corophium volutator (estuarine amphipod)

These were tested according to the chronic protocol described in Section 2 (see Appendix B) although a 30 day exposure duration was used in order to maintain comparability with the other selected chronic test methods. The test animals were unfed.

Arenicola marina (estuarine polychaete)

This test species has already been subject to a Paris Commission ring test and has been identified by the Environment Agency as a potentially suitable test species for regulatory purposes. The method selected is a ten day test over which the sublethal endpoint of cast formation is measured on a daily basis, and mortality is assessed at the end of the exposure period. Animals were supplied by a Mr R Pipe of Burnham on Crouch, Essex on 18 May 1995. These animals were graded into three sizes. The smaller animals were used for the study. The test animals were unfed throughout the exposure duration.

Neanthes areneceodentata (estuarine polychaete)

This test organism has been used in the United States (Reish 1980) and a test protocol is currently being developed by MAFF. The exposure duration is 30 days after which mortality and differences in growth, measured as dry weight, are assessed. Juvenile animals were supplied for this test by Dr D Reish of California State University on 17 May 1995. The test animals had to be fed over the test duration to ensure survival.

Macoma balthica (estuarine bivalve)

This test species has been use previously at WRc to assess bioaccumulation of metals from sediments. It was included here to represent a third taxonomic group of organisms. The animals were exposed for 30 days after which mortality was assessed. Adults were collected from a field site in Teignmouth, Devon on 29 April 1995. The test animals had to fed over the test duration to ensure survival.

4.2.2 Sediments

The sediment used in this study was Sediment 1 used in Section 2. This was adjusted to give three sediment types so that comparisons of control survival across sediments could be made.

Sediment was collected, on 14 March 1995. This was passed through a 500 μm sieve before use. Three sediments of different particle sizes (fine, medium and coarse) were prepared by blending differing proportions of Poole sediment with BDH acid washed sand. Table 4.1 shows the particle size distribution, total organic carbon (TOC) and percentage solids content for the three sediment types.

Table 4.1 Percentage particle size distribution and TOC for the three sediment types

	Fine	Medium	Coarse
Particle size (%)			
2000 - 600 μm	1	0	0
600 - 212 μm	1	24	26
212 - 63 μm	1	34	67
63 - 20 μm	13	6	1
20 - 2 μm	46	18	3
<2 μm	38	18	3
TOC mg l^{-1} (%)	2880 (0.87)	2980 (0.90)	1420 (0.47)
Solids (%)	33	33	30

These three sediments were then spiked with lindane to produce the following nominal concentrations:

0.0 (control), 180, 320, 560, 1000 and 1800 ng g^{-1} (dry weight).

Spiking was undertaken using an industrial food mixer (Crypto Peerless). Solutions of lindane were prepared in acetone such that the addition of a common volume of each solution to a common weight of sediment would produce the required nominal concentration of lindane in the sediment. Known weights of sediment were then added to a silanised stainless steel

mixing bowl, the mixer started and the requisite volumes of lindane in acetone added slowly while mixing the sediment. Mixing was continued for 1 hour for each batch of spiked sediment. Control sediment was spiked with the same volume of acetone as the treatments containing lindane (0.2 ml kg^{-1} wet weight).

The concentrations of lindane achieved in the spiked sediments were analysed by extracting known weights of material with hexane, followed by measurement on a Varian 3500 gas chromatograph with electron capture detection. The coarse sediment presented handling problems as regards maintenance of homogeneity during sampling and during addition of the sediment to the test vessels. The high proportion of sand to sediment meant that settlement of sand grains occurred rapidly with this material once agitation ceased. Thus, the taking of representative samples for analysis proved difficult. In order to examine the implications of this, a large sample (500 g) was taken from the highest concentration and analysed to investigate partitioning of the lindane between the sand, fine sediment and water fractions present in the sample.

4.2.3 Test methods

Table 4.2 details the test vessels and volumes used for each toxicity test. The tests were carried out simultaneously at a constant temperature of 15 ± 1 °C.

Table 4.2 Test vessels and volumes used for the toxicity tests

Test organism	Vessel volume (l)	Weight of wet sediment (g)	Volume of overlying water (ml)	No. of replicates	No. of animals
Arenicola	2	750	600	3	5
Macoma	2	750	600	3	5
Neanthes	2	750	600	2	5
Corophium	1	60	400	2	10

Arenicola tests were conducted over 10 days, using survival and cast formation as endpoints. *Corophium, Macoma* and *Neanthes* tests were conducted over 30 days with survival as the end point in all cases,. In addition, growth (determined from dry weight) was also examined with *Corophium* and *Neanthes*. Water quality data were recorded at the start and end of each test.

4.3 Results

Table 4.3 summarises the water quality data of each test which were within acceptable ranges as specified in the protocols (see Appendix B).

Table 4.3 Summary water quality data for the tests with lindane and three sediment types

	Temperature (°C)	pH	Dissolved oxygen (% ASV)	Salinity (‰)
Fine				
Arenicola	14.6 - 15.0	7.69 - 7.94	96 - 100	36 - 37
Corophium	14.4 - 14.6	7.85 - 8.00	88 - 100	34 - 36
Macoma	14.4 - 14.6	7.65 - 7.95	89 - 99	38 - 39
Neanthes	14.4 - 14.6	7.79 - 8.11	91 - 100	34 - 35
Medium				
Arenicola	14.6 - 15.1	7.71 - 7.84	94 - 100	36 - 38
Corophium	14.4 - 14.6	7.74 - 7.89	87 - 100	36 - 38
Macoma	14.4 - 14.6	7.51 - 7.79	86 - 100	40 - 41
Neanthes	14.4 - 14.6	7.80 - 8.00	86 - 100	35 - 36
Coarse				
Arenicola	14.6 - 14.9	7.75 - 7.85	96 - 100	36- 38
Corophium	14.4 - 14.6	7.84 - 7.98	88 - 100	36 - 37
Macoma	14.4 - 14.6	7.56 - 7.80	84 - 100	40 - 42
Neanthes	14.4 - 14.6	7.77 - 7.81	89 - 100	36 - 37

4.3.1 Chemical analysis

Table 4.4 shows the exposure concentrations achieved by the spiking method used. Concentrations have been normalised for organic carbon using the method of Di Toro *et al.* (1991):

$$\text{organic carbon normalised concentration} = \frac{\text{total sediment concentration}}{f_{oc}}$$

where f_{oc} = mass fraction of organic carbon

Table 4.4 Nominal vs actual concentrations for the three sediment types. Numbers in parenthesis represent lindane concentrations normalised for organic carbon content (ng g^{-1} C)

Nominal concentration (ng g^{-1})	Sediment type		
	Fine	Medium	Coarse
0 (control)	0	0	0
180	190 (218)	200 (221)	410 (866)
320	240 (275)	320 (354)	1710 (3613)
560	380 (435)	540 (598)	1050 (2218)
1000	760 (870)	970 (1074)	3920 (8282)
1800	1360 (1558)	1840 (2038)	5040 (10648)

The analysis of the sample of coarse sediment for lindane partitioning showed that lindane was sequestered greatest by the fine sediment fraction of the sample (18.7 $\mu g\, g^{-1}$) and much less so by the sand (0.38 $\mu g\, g^{-1}$). Very little lindane (0.14 $\mu g\, ml^{-1}$) remained in solution. These data show that the problems experienced in achieving a homogenous sample influenced the results of the analysis. It proved impossible to sample this sediment and assure that the relative proportions of fines, sand and water were maintained in the sample. The same problem was experienced when adding sediment to the test vessels. Hence, for the coarse sediment it is not clear that the test organisms were exposed to these analysed concentrations, as each vessel contained a slightly different ratio of fines to sand.

4.3.2 Toxicity of lindane to four species of marine/estuarine animals

Table 4.5 shows the toxicity data from tests with all four species in the three sediment types.

Table 4.6 shows the data for LC/EC_{50}s and lowest observed effect concentrations (LOECs) for lindane to the test species, normalised for organic carbon content.

Table 4.5 Results from toxicity tests with lindane in three sediment types

Concentration (ng g^{-1} C)	*Arenicola*		*Corophium*			*Macoma*	*Neanthes*		
	Total casts	% survival	Total biomass (mg)	Mean individual dry weight (mg)	% survival	% survival	Total biomass	Mean individual dry weight (mg)	% survival
Fine									
0	99	87	10.51	0.66	80	87	6.47	0.65	100
218	101	87	8.68	0.58	75	93	4.83	0.54	90
275	107	100	8.34	0.52	80	100	7.11	0.89	90
435	77	87	7.44	0.57	65	100	5.78	0.83	90
870	74	100	6.61	0.60	55	100	3.84	0.64	60
1558	64**	100	2.09**	0.70	15**	93	4.57	0.51	90
Medium									
0	123	93	4.39#	1.10	40#	100	3.59	0.40	90
221	111	93	7.14	1.23	30	100	2.70	0.30	90
354	92	87	6.04	1.21	25	93	3.05	0.38	80
598	89	100	4.54	1.14	20	80	3.40	0.49	70
1074	64**	73	1.93	1.93	5	93	3.11	0.39	80
2038	39**	100	0.00	0.00	0	87	3.60	0.40	90
Coarse									

Concentration (ng g^{-1} C)	*Arenicola*		*Corophium*			*Macoma*	*Neanthes*		
	Total casts	% survival	Total biomass (mg)	Mean individual dry weight (mg)	% survival	% survival	Total biomass	Mean individual dry weight (mg)	% survival
0	79	93	7.37	0.61	80	87	2.33	0.33	70
866	41	93	0.52	0.14	25	73	1.68	0.28	60
3613	47	93	0.00	0.00	0	93	2.59	0.86	30
2218	43	93	0.00	0.00	0	80	1.08	0.22	50
8282	103	100	0.00	0.00	0	93	1.92	0.27	70
10648	24	100	0.00	0.00	0	93	1.91	0.17	70

data from one replicate
** significant at p = 0.05

Table 4.6 EC_{50}s and LOECs for lindane in three sediment types to the test species (ng g^{-1} C)

Test species	Fine	Medium	Coarse
Arenicola			
10-day EC_{50} (cast formation)	2120 (985 - 4560)* [1850 (860 - 3980)]#	1141 (598 - 2159) [1030 (540 - 1950)]	no dose related response
10-day LOEC (cast formation)	1558 [1360]	1074 [970]	no dose related response
Corophium			
30-day LC_{50} (mortality)	985 (733 - 1329) [860 (640 - 1160)]	277 (155- 498) [250 (140 - 450)]	100% mortality at 3613 ng g^{-1} C and above
30-day LOEC (dry weight-growth)	1558 [1360]	1074 [970]	-
Macoma	no observed effect	no observed effect	no observed effect
Neanthes	no observed effect	no observed effect	no observed effect

* numbers in ( ) are 95% confidence intervals
\# concentrations in [ ] are not normalised for organic carbon, units are ng g^{-1} dry sediment.

4.4 Discussion

4.4.1 Sediment spiking

The method used for spiking the sediments was both accurate and rapid. Actual concentrations were close to nominals for both fine and medium grained sediments. However, the coarse grained sediment presented problems of homogeneity due to the rapid settlement of sand grains from the finer sediment and water phases. Analysis showed that the lindane partitioned preferentially to the fine sediment material, with very little binding to the sand. Hence, we could not ensure that sediment subsamples placed in the test vessels, or the samples of this sediment taken for chemical analysis, were representative of the proportions of the two solid fractions present in the body of sediment as a whole. Thus, the concentration of lindane was dependent on the ratio of sand to sediment present in each vessel and as such would not be the same in 'replicate' vessels. Therefore, the results obtained with the coarse sediment, both analytical and toxicological, must be treated with caution. The other two sediments were not affected by this problem as the higher proportions of fine material present kept the sand grains suspended in the body of sediment, preventing their rapid settlement and preserving

homogeneity of the sediment while vessels were filled or analytical samples taken. The problem of preserving sample homogeneity may also arise when environmental sediments of a similar composition to the coarse sediment used for this work are used for assessment of their toxicity.

4.4.2 Water quality

Measurements of water quality showed that the parameters monitored remained within acceptable limits on the occasions when measurements were made. The limited number of measurements minimised undue disturbance of the sediment during exposure of the test organisms, reducing stress on the test system.

4.4.3 Toxic responses

Arenicola, *Macoma* and *Corophium* survival and growth appeared not to be affected by the physical nature (coarseness and organic carbon content) of the sediment. Control survival was >80% in all three sediment types. *Neanthes* control survival declined as the sediment structure became coarser, indicating that this species was adversely affected by the coarse grained sediment used.

No clear dose related response was seen with any of the test species in the coarse sediment. This may have been due to problems with sediment sampling and heterogeneity as discussed above. However, with *Corophium volutator*, 100% mortality was seen at an organic carbon content normalised lindane concentration of 3612 ng g^{-1} and above. This is in agreement with the toxicity data of the fine and medium sediments. For example, in the medium sediment 100% mortality was seen at 2038 ng g^{-1} and so at the concentrations found in the coarse sediment, mortality would be expected.

Of the four tests, the *C. volutator* was the most sensitive to lindane in terms of lethality. A reduction in survival was seen in all three sediment types as lindane concentration increased. No reduction in survival was seen with any of the other test animals. Lethality of *Macoma, Neanthes* and *Arenicola* should not be considered for regulatory purposes.

In terms of sublethal endpoints, *C. volutator* and *A. marina* exhibited similar sensitivities for both fine and medium sediments, for growth and cast formation respectively. The Lowest Observed Effect Concentration (LOEC) for these sublethal endpoints was the same for both test species. However, LOECs for both endpoints were significantly higher than the 30-day LC_{50} values for *Corophium*. This demonstrates that the lethal endpoint for *Corophium* is the most sensitive measure of long-term exposure to lindane. The additional sensitivity of the chronic exposure compared with the acute exposure should be compared more extensively before the optimal exposure duration is determined as discussed in Sections 2 and 3.

The *Arenicola* test was not as sensitive as *Corophium* test even when sublethal endpoints were measured. Another disadvantage with this test species is that it became clear as the experiment proceeded that the casts produced by animals in the coarse sediment were smaller than those produced by animals in the medium and fine sediments. The size differential was not related to the size of the animals used and supported the total cast number data which showed that less feeding was occurring in the coarse sediment than the others. However, rather than

causing an overall reduction in total cast number, this introduced more between-treatment variation. Control survival and cast formation was still acceptable and it is possible that the discrepancies were due to sediment heterogeneity. The advantage of using the *Arenicola* test is that it employs a 10-day exposure as opposed to the 30-day *Corophium* exposure. This may be beneficial in terms of cost-benefit. Therefore further comparisons of *Arenicola* cast-formation and *Corophium* lethality should be investigated for a range of field sediments.

5. CONCLUSIONS

5.1 Acute versus chronic tests

5.1.1 *Chironomus riparius*

The *C. riparius* chronic test was marginally more sensitive than the ten-day acute tests performed previously with lindane-spiked sediment. However, the difference in sensitivity between the two exposure durations, as determined by lethality, was much smaller than expected. Development to fourth instar was the most reliable and sensitive sublethal endpoint used but significant differences were only observed at lindane and permethrin concentrations also causing significant differences in mortality. The fact that the chronic lethal and sublethal toxicity was less than expected could either reflect the mode of toxic action of the test chemicals or demonstrate that the chronic test is no more sensitive than the acute test, possibly because the former requires additional feeding which may reduce exposure. Comparative tests are required for a wider range of contaminant types before the relative sensitivity of the chronic test can be determined. No additional sensitivity was demonstrated in the chronic test with field sediments compared to the acute tests but this was due to the non-toxic nature of the sediments after storage.

The OECD have recommended chronic tests for risk assessment of pesticides general chemicals in sediments in order that all routes of exposure for different chemical types can be included in the test duration. In the OECD forum, various test methods have been considered including the protocol developed in this programme. The selection of an OECD method was based on the current status of existing guidelines in Europe, Canada and North America, and different aspects of each test are incorporated into the final draft protocol which should be available during 1997. The test guideline will be for a chronic test initiated with first instar larvae measuring adult emergence as the primary endpoint. The test system is essentially the same as that developed here except the animals are fed 0.5 mg of ground tetramin fish food per larvae per day for the first ten days, and 1 mg per larvae for the remainder of the test. In a recent study performed for the DoE, the OECD protocol was compared with the protocol developed in this programme for permethrin-spiked sediment (Fleming *et al.* 1997). The OECD protocol was more sensitive suggesting that inclusion of emergence, despite the additional feeding, enhance the test when used for sediments spiked with permethrin.

The OECD method has been developed for assessment chemicals spiked into sediments. It has not been used widely as a bioassay. One difficulty with using emergence as an endpoint in bioassays is that it is influenced by food availability (van de Gucthe, pers. comm. 1995). As different field sediments have different organic matter type and content, they will also have varying levels of food availability. This may lead to differences in emergence which are unrelated to toxicity which has implications for the use of the OECD guideline in assays of field sediments. This may be normalised to some extent by the large additional food ration supplied in this test.

In order to determine a suitable exposure duration, life stage and measurement endpoint for a chronic *C. riparius* test for Environment Agency purposes, a small amount of further work is required to determine which of the following tests can best discriminate between field sediments from contamination gradients:

- 10 day acute test measuring lethality (fed and unfed);
- 28 day chronic test measuring larval mortality and development (as developed in this programme); and
- OECD emergence test.

A workshop funded by the European Commission was held in Pallanza, Italy in 1996, to discuss the status of Chironomidae guidelines for use in pesticide risk assessment. This has been supplied to the Environment Agency under separate cover.

5.1.2 *Corophium volutator*

The *Corophium volutator* chronic test was more sensitive that acute tests with lindane in terms of the lethality endpoint. The additional sensitivity of this species over long-term exposure durations compared with *C. riparius* could be explained by the fact that no additional feeding was given. However, this difference in sensitivity was not demonstrated by the field sediment trials even when the sublethal growth endpoint was measured. This could be due to the mode of action of the contaminants in the field sediments, a decrease in toxicity during storage, or that the chronic test is no more sensitive than the acute test to field sediments.

The measurement of sublethal endpoints is a problem for this test species. The dry weight endpoint was shown to be unreliable and insensitive in the spiked-sediment chronic tests with lindane and permethrin. When decreases in dry weight were observed with increasing concentration. Statistically significant differences were rarely seen due to variability between replicates. The animals were accurately sized on test termination and so a reduction in variability could only be potentially be reduced by increasing the number of replicates. this would have cost and resource implications. An additional difficulty with the dry weight endpoint is that growth tended to increase as mortality increased, suggesting that the animals were competing for food. In the field sediment tests it was more discriminatory over a ten day test duration. This could be partially explained by the food availability issue. In the spiked sediment test with lindane, significant differences were seen in mean individual dry weights between treatments in one of the two sediment types used in which growth rate in control vessels was highest. This was more likely to be due to organic matter type as the organic carbon content was less in the sediment in which increased growth was seen. For the field sediment tests, food availability would have been less limiting over a ten day duration than 28 days. Therefore, the utility of the dry weight endpoint may be due to food availability. For spiked sediment tests, enhancing this endpoint would simply mean selecting a reference sediment in which optimum growth over a long-term period could be demonstrated. For field sediments, the growth endpoint could be enhanced by additional feeding. Further research is

required to compare the relative sensitivity of acute and chronic tests for field sediments, focusing on the need for additional feeding. Feeding could increase test discrimination of the chronic test compared with the acute test, enhance the dry weight endpoint and it could be useful in normalising between sediment types which will have different levels of food available.

5.2 Reproducibility

In inter-laboratory comparisons, survival in both of the *C. volutator* and the *C. riparius* tests was found to be the least variable endpoint measured for both laboratory-spiked sediments and field contaminated sediments. For *C. riparius,* coefficients of variation between laboratories for all endpoints measured were lowest for the tests with field sediments. However, this was reflective of the fact that no effects were observed in these tests. For the spiked-sediment tests the *C. riparius* did show variability for all endpoints. This demonstrates the need for experienced personnel when conducting chronic tests when compared to the acute tests developed previously (Fleming *et al.* 1994). For the *C. volutator* tests, the inter-laboratory comparison with permethrin-spiked sediments showed a high degree of reproducibility for the survival endpoint. In the field contaminated sediment tests with this species, a high degree of reproducibility was seen between the two laboratories with most experience of the test method. Again, this demonstrates the need for experience personnel when conducting these tests.

Much of the variability in the spiked sediment tests could be attributed to the spiking technique and sample handling in the different laboratories. In the lindane tests, samples for each laboratory were spiked individually. In the permethrin test, a batch spiking process was used and the variability in the toxicity test results was reduced. Because sediment-spiking is a more complex process than delivery of test substance to water-only tests, an additional source of variability is encountered with sediment tests.

5.3 Comparison of four marine species

A comparison of four macrobenthic sediment test organisms indicated that *Corophium volutator* was the most sensitive (when exposed to lindane) and the most tolerant of differences in sediment type. Lethality of *Corophium* over a chronic test duration and the sub-lethal endpoint of cast-formation in *Arenicola* over ten days were of similar sensitivity. However, the latter was more influenced by sediment type, showing a higher degree of variability on coarse sediment.

5.4 Meiofaunal polychaetes

Ophryotrocha spp was not as sensitive to the field sediments tested here as *Corophium volutator.* Although these species do have advantages over the macrofaunal sediment tests developed in this programme in terms of laboratory space and sediment sample sizes, more research is required on the relative sensitivities of the two to a range of contaminants and contaminated field sediments before being considered for regulatory purposes.

6. RECOMMENDATIONS

1. Chronic tests have been recommended by the OECD for risk assessment of sediment-bound contaminants because they are more likely to simulate realistic exposure scenarios in the field than acute tests. Chronic tests are particularly important for general quality assessment because they may allow additional discrimination between sites containing low to medium levels of contaminants. However, chronic tests should only be used in place of established acute tests if they can be shown to produce better information. Chronic tests are more difficult to perform, are potentially less reproducible, and require a longer test duration.

2. Both acute and chronic tests developed here for *Chironomus riparius* were included in the GQA pilot study performed in 1996. The data from this study are still being analysed to determine the relative sensitivity of the two. If the chronic test is shown to be less cost-effective in terms of reliability and discrimination, a limited number of field sediment tests should be conducted to compare the sensitivity of the acute test and the recently developed OECD protocol based on emergence. This should be performed before the GQA scheme is widely implemented.

3. Both acute and chronic tests developed here for *Corophium volutator* were used in the GQA pilot study. Based on the recommendations made in this programme, the chronic test was supplied with additional food. If results show the acute tests to be more cost-effective, a limited number of tests should be performed with field sediments to compare the effects of feeding and non-feeding on discrimination of the chronic tests. Measurement of sublethal toxicity for this species is a problem. Behavioural responses such as sediment avoidance have been measured previously (Davies *et al.* 1991) but this type of response is difficult to calibrate against population effects. Reproductive endpoints such as the number of offspring produced or the time to reproduction could be investigated further although this would mean trying to 'speed up' the life cycle. This could potentially achieved by increasing the test temperature which has been done for chronic tests with the North American species *Leptocheirus plumulosus* (Dewitt, MSL, USA, pers. com. 1996) but this does reduce ecological relevance and may have implications for sediment physico-chemical character, quality and contaminant partitioning. Alternative endpoints for *C. volutator* have not been investigated in this project.

4. *Arenicola marina* tests have been included in the GQA pilot study as the sublethal endpoint of cast formation may provide additional information on sublethal toxicity in conjunction with *Corophium volutator* tests. The two species may provide different information for different sediment and contaminant types. However, *A. marina* does seem to be sensitive to differences in grain size and may not be useful for the widespread application needed for GQA purposes. If the pilot study shows that *Arenicola marina* provides no additional information compared with *Corophium volutator*, *Corophium* should be used in preference.

5. The meiofaunal polychaetes require further research to determine their relative sensitivity to other test species before being considered for regulatory purposes.

REFERENCES

Bryant, V., McLusky, D.S., Roddie, K. and Newberry, D.M. (1984) Effect of temperature and salinity on the toxicity of chromium to three estuarine invertebrates (*Corophium volutator, Macoma balthica, Nereis diversicolor*), *Marine Ecology Progress Series*, **20**, 137-149.

Bryant, V., McLusky, D.S., Campbell, R. and Newberry, D.M. (1985) Effect of temperature and salinity on the toxicity of arsenic to three estuarine invertebrates (*Corophium volutator, Macoma balthica, Tubifex costatus*), *Marine Ecology Progress Series*, **24**, 129-137.

Burton, G.A. Jr (1991) Assessing the toxicity of freshwater sediments, *Environmental Toxicology and Chemistry,* **10**, 1585-1627.

Burton, G.A. Jr, Nelson, M.K. and Ingersoll, C.G. (1992) Freshwater Benthic Toxicity Tests, In: Ed: G.A. Burton, Jr, Sediment Toxicity Assessment, Lewis Publishers.

Burton, G.A. Jr and MacPherson, C. (1995) Sediment Toxicity Testing, In: Eds: D.J. Hoffman, B.A. Ratner, G.A. Burton Jr and J. Cairns Jr, *Handbook of Ecotoxicology.*

Caron, G., Suffet, I.H. and Belton, T. (1985) Effect of dissolved organic carbon on the environmental distribution of nonpolar organic compounds, *Chemosphere*, **14**(8), 993-1000.

Ciarelli, S., Vonck, W.A.P.M.A. and van Straalen, N.M. (1997) Reproducibility of spiked-sediment bioassays using the marine benthic amphipod, *Corophium Volutator, Marine Environmental Research*, **4**, 329-343.

Davies, R., Roddie, B. and Jerman, E. (1991) The development of chemical and biological methods for assessing the toxicity of sediment-bound contaminants, WRc NR 2729/1.

Dewitt, T.H., Redmond, M.S., Sewall, J.E. and Swartz, R.C. (1992) Development of a chronic sediment toxicity test for marine benthic amphipods. US EPA document for the Chesapeake Bay Progam.

Di Toro, D.M., Zarba, C.S., Hansen, D.J., Berry, W.J., Swartz, R.C., Cowan, C.E., Pavlou, S.P., Allen, H.E., Thomas, N.A. and Paquin, P.R. (1991) Technical basis for establishing sediment quality criteria for non-ionic organic chemicals using equilibrium partitioning, *Environmental Toxicology and Chemistry*, **10**, 1541-1583.

Erdem, C. and Meadows, P.S. (1980) The influence of mercury on the burrowing behaviour of *Corophium volutator, Marine Biology*, **56**, 233-237.

Fleming, R., Grootelaar, E.M., Van de Gutche, C., Ciarelli, S., Borchert, J., Looise, B., Guerra, M., Gaudencio, O., Castro, L., Karbe, L. and Vonck, W. (1996) Validation of sediment toxicity tests for poorly soluble substances. WRc report to the European Commission EV5V-CT91-0405.

Fleming, R., Hermsen, W., Spence, S., Vine, S. and Crane, M. (1993) The effects of partitioning on toxicity, WRc TR report, 017/14/W.

Ingersoll, C.G. and Nelson, M.K. (1990) Testing sediment toxicity with *Hyalella azteca* and *Chironomus riparius*, Aquatic Toxicology and Risk Assessment, Thirteenth Volume ASTM STP 1096, Eds: W.G. Landis and W.H. van der Schalie, American Society for Testing and Materials, Philadelphia.

Kukkonen, J. and Landrum, P.F. (1994) Toxicokinetics and toxicity of sediment-associated pyrene to *Lumbriculus variegatus* (Oligochaeta), *Environmental Toxicology and Chemistry*, **13**, 1457-1468.

Lee, W.Y. (1977) Some laboratory cultured crustaceans for marine pollution studies, *Marine Pollution Bulletin*, **8**, 258-259.

Matthiessen, P. and Thain, J.E. (1989) A method for studying the impact of polluted marine sediments on intertidal colonising organisms; tests with diesel-based drilling mud and tributyltin antifouling paint, *Hydrobiologia*, **188/189**, 477-485.

Meadows, P.S. (1964) Experiments on substrate selection by *Corophium volutator* (pallas): depth selection and population density, *Journal of Experimental Biology*, **41**, 677-687.

Meadows, P.S. and Reid, P. (1964) Experiments on substrate selection by *Corophium* sps.: films and bacteria on sand particles, *Journal of Experimental Biology,* **41**, 499-511.

Meadows, P.S., Deans, E.A. and Anderson, J.G. (1981) Responses of *Corophium volutator* to sediment sulphide, U *Journal of the Marine Biological Association UK*, **61**, 739-748.

Mills, A. and Fish, J.D. (1980) Effects of salinity and temperature on *Corophium volutator* and *Corophium arenarium* (Crustacea: Amphipoda), with particular reference to distribution, *Marine Biology*, **58**, 153-161.

Nebeker, A.V., Cairns, M.A. and Wise, C.M. (1984a) Relative sensitivity of *Chironomus tentans* life stages to copper, *Environmental Toxicology and Chemistry*, **3**, 151-158.

Nebeker, A.V., Cairns, M.A., Gakstatter, J.H., Malueg, K.W., Schuytema, G.S. and Krawczyk, D.F. (1984b) Biological methods for determining toxicity of contaminated freshwater sediments to invertebrates, *Environmental Toxicology and Chemistry*, **3**, 617-630.

OECD (1995) Detailed review paper: aquatic testing methods for pesticides and industrial chemicals. Prepared for the National Co-ordinators of the OECD Test Guidelines Programme by Water Quality Institute, Denmark and RIVM, the Netherlands. January 1995.

Plesha, D.P., Stein, J.E., Schiewe, M.H., McCaine, B.B. and Varanasi, V. (1988) Toxicity of marine sediments supplemented with mixtures of selected chlorinated and aromatic hydrocarbons to the infaunal amphipod *Rhepoxynius abronius*, *Marine Environmental Research*, **25**, 85-97.

Reichart, W.L., Eberhart Le. B. and Varanasi, V. (1985) Exposure of two species of deposit-feeding amphipods to sediment-associated 3[H]benzo(a)pyrene: uptake, matabolim and covalent binding to tissue macromolecules, *Aquatic Toxicology*, **6**, 45-56.

Roddie, B.D., Kedwards, T.J. and Davies, R.E. (1991) An investigation of the toxicity of beach sand contaminated by spillage of crude oil. WRc report NR 2649.

Saleh, F.Y., Dickson, K.L. and Rodgers, J.H. Jr (1982) Fate of lindane in the aquatic environment: rate constants of physical and chemical processes, *Environmental Toxicology and Chemistry*, **1**, 289-297.

Sims, I. and Fleming, R. (1995) Comparison of four estuarine sediment toxicity test species, poster presentation to SETAC World Congress, Vancouver, 1995.

Wentsel, R., McIntosh, A., McCafferty, W.P., Atchison, G. and Anderson, V. (1977) Avoidance response of midge larvae (*Chironomus tentans*) to sediments containing heavy metals, *Hydrobiologia*, **55**,171-175.

APPENDIX A

MEIOFAUNAL POLYCHAETES

DEVELOPMENT OF SEDIMENT TOXICITY TESTS AND BIOASSAYS WITH MARINE MEIOFAUNAL POLYCHAETES

Mark Crane and Barry Shepherd

Division of Biology
School of Biological Sciences
Royal Holloway University of London
Egham
Surrey
TW20 OEX

Final Report
October 1996

1. INTRODUCTION

1.1 The ecotoxicological importance of sediments

It is known that freshwater and marine sediments sequester certain xenobiotics (Knezovich *et al.*, 1987; Luoma and Carter, 1993; Chandler *et al.*, 1994). Suspended particles adsorb the xenobiotics (especially if they are hydrophobic or metallic) from the water phase and, when they settle, remove them from the water column (Hill *et al.*, 1993). Thus, sediments act as a sink for many pollutants (Chandler *et al.*, 1994). Marine pollutants originate from many sources; direct discharge occurs as well as indirect discharge from rivers and the atmosphere. Because of the dynamics of suspended particles, the smaller particles are transported to estuaries and the sea (Selley, 1982). In general, the physico-chemical properties of smaller particles allow more contaminants to be carried for a given volume than do large particles (Selley, 1982). For this reason substantial proportions of freshwater contaminants find their way into the marine environment.

The many thousands of different compounds humans have released into the environment have created complex mixtures within sediments (Giesy and Hoke, 1989). Many of these chemicals are rendered harmless by equally complex reactions. However, some can become more toxic when acting in concert with other chemicals (Ehrenstrom, 1979; Cairns, 1986). Such contamination can result in a direct or indirect effect on infaunal and epifaunal organisms. Fowler *et al.* (1978) demonstrated that up to 99% of an organism's body-burden of chemicals can be derived from the sediments rather than the surrounding water. Additionally, at times of re-suspension, the contaminated sediments can release the xenobiotics back into the water column (Burgess *et al.*, 1993), or the sediments can be consumed by filter-feeding organisms with subsequent partitioning of sorbed contaminants into these organisms (Hermsen *et al.*, 1994). Unfortunately, the majority of this activity occurs in estuarine and near-coastal regions. These regions are highly valuable for the feeding and spawning of commercial fish (Knox, 1977; Knezovich *et al.*, 1987; Burgess *et al.*, 1993). They are also the sites of aquaculture, recreation and where most of the ocean's biodiversity exists (Dubinsky, 1990).

1.2 Toxicity test and bioassay systems for marine sediments

2. the responses of control organisms should be predictable and constant;

3. the assay organisms should respond similarly to many classes of toxicants, or a battery of test organisms sought, with sufficiently different responses to test most chemical groups;

4. the results of the assays should be related to ecologically relevant processes under field conditions;

5. the results of bioassays should be related to sediment or water quality standards and criteria;

6. the bioassays should be applicable to a number of sediment types and environments;

7. the assays and dilution methods should be chosen to provide information, that is correlated with observed adverse effects on organisms under field conditions;

8. the assays should be rapid, reliable, inexpensive, and easily implemented .

9. the assays should be standardized to facilitate widespread use.

10. the assays should be sensitive enough to identify potential problem sediments yet discriminatory enough to permit ranking of the relative toxicity of many samples.

These criteria map well onto those proposed by the Environment Agency for the selection of test organisms within the Direct Toxicity Assessment programme (Environment Agency, 1996).

But why should organisms be used instead of just chemical and physical analysis? Cairns (1986) notes that toxicity is not a measurable characteristic of a chemical. Also, in the case of sediments and mixtures of chemicals, scientists do not fully understand the processes that occur (Luoma and Carter, 1993), and therefore may not be able to predict toxicity

sublethal effects occurred at much lower concentrations, with 5-6 week EC50 values for reproduction of 0.025 mg l^{-1} (PCP), 0.01 mg l^{-1} (DCA) and 0.0009 mg l^{-1} (dieldrin). Parker (1984) showed in a series of 48-h toxicity tests that *O.diadema* is of similar acute sensitivity to contaminants as other polychaetes. Knowles and Greenwood (1994) also demonstrated the usefulness of this species in assessing the effects of radioactive waste on reproductive output.

Other *Ophryotrocha* species have also been used in water-only toxicity tests. *O. labronica* was tested by Reish and Lemay (1991), who obtained the following 95% confidence intervals for the 96-h LC50: 1.0-2.1 mg As l^{-1}, 3.3-4.2 mg Cd l^{-1}, 1.9-5.5 mg Cr l^{-1}, 0.03-0.06 mg Cu l^{-1}, 0.1-0.23 mg Hg l^{-1} and 1.36-2.3 mg Zn l^{-1}. Several species of *Ophryotrocha* exposed to phenol had 96-h LC50 values of between 110-350 mg l^{-1}.

Ophryotrocha are brooders and their larvae will remain in, or close to the sediment. Thus, by completing their whole life-cycle in the sediment they are probably more relevant organisms for use in sediment tests than are organisms with planktotrophic larvae. Eggs and larvae that form part of the plankton escape the effects of contamination during what may be their most sensitive phase. Much is known about *Ophryotrocha* as culture and test organisms, but no trials have been conducted to assess their environmental parameter preferences in sediments

1.4 Project objectives

Sediment characteristics have a profound effect on the bioavailability of contaminants (Knezovich *et al.*, 1987; Giesy and Hoke, 1989). This is primarily due to the bonds between the particles and contaminants. Particle size and organic matter content are particularly important in controlling bioavailability (Suedel and Roberts, 1994), and can also have a direct effect on the test organism (Hill *et al.*, 1993). Therefore, an organism's response to such characteristics must be known prior to its use in sediment bioassays. Investigation of these aspects must be the first stage in establishing the potential usefulness of a test organism.

2. MATERIALS AND METHODS

2.1 Test organisms

The three species of *Ophryotrocha* used in this study have been studied previously by Professor Bertil Åkesson at the University of Goteborg, Sweden, and found to be suitable for culture in the laboratory. The history of each culture used in this study is known. *O. p. siberti* and *O. hartmanni* were obtained from Dr John Knowles at the Ministry of Agriculture, Fisheries and Food (MAFF), Lowestoft. These cultures originated from wild individuals collected at Oban (UK). *O. diadema* were originally received at Royal Holloway from Bertil Åkesson, but were later supplemented with organisms from John Knowles (also originally obtained from Åkesson). Åkesson obtained his original culture of *O. diadema* from Los Angeles harbour, California. Having been kept in culture for many years all three species of *Ophryotrocha* are now genetically inbred and relatively homogeneous. This should provide low variation in response, and more precise results. However, it is inevitable that cultures set up from only a few individuals can suffer genetic drift or bottle-necking (Perkins, 1972; Skelton, 1993).

Various aspects of the species' autecology are shown in Table 1. Where shown, the figures have been obtained from the author cited. The other figures were obtained from the present study and correspond favourably with those obtained by Åkesson (1973, 1976) and Tennant (1984).

	O. hartmanni	*O. p. siberti*	*O. diadema*
optimum temp. °C	15	17	20
sexual strategy (Åkesson)	sequential protandrous hermaphrodite.	sequential protandrous herm. with ability to reverse.	simultaneous protandrous herm.
maximum length. mm	15	15	5
male size: segment No (Åkesson)	6-19 segments	9-17 segments	6-12 segments
Age to male: days	13 days	14 days	10 days
fem.size: segment No (Åkesson)	19	20	
age to female: days	32 days	28 days	23 days
average number of eggs	20-45 eggs	250 eggs	29 eggs

2.2 Procedure to obtain cohorts of organisms for use in an experiment.

When the chosen endpoint of an experiment is reproduction all individuals of the same sex should be in the same reproductive stage and condition (Åkesson, 1970; 1976; *pers. comm.*). When reproductive output is age-dependent, as is the case with these three species (Åkesson, 1976; Tennant, 1984), cohorts are even more desirable. With sequential hermaphrodites, cohorts staggered in age by the difference between one sex and the other are required. Different ages would introduce another variable which would then increase variation, and thus reduce the power of any statistical analysis. This is particularly important for *O. hartmanni*. If the males in one treatment group were several days older than those in other groups, there is the possibility that male gametes would be absent when the females have mature oocytes. Conversely, some males may mature to female, thus adding further reproductive output to some treatments and not to others. The opposite is true for *O. diadema* in which male and female gametes can be produced by one individual at the same time (Åkesson, 1976; Tennant, 1984).

The following is the procedure that should be adopted to produce two cohorts in *O. hartmanni*.

1. Isolate a number of non-gravid females for 7 days at 2 - 5 °C less than normal temperature (cohort 1. females).

2. Unite above females with males and return to normal temperature.

3. Isolate a number of non-gravid females for 7 days at 2-5 °C less than normal temperature (cohort 2. males) 12 days after the first set.

4. Unite second set of females with males and return to normal temperature.

5. Remove all egg masses not laid in first 3 days. Adults must remain to tend egg masses (Åkesson, 1973).

6. When larvae hatch and become free-swimming, adults must be removed.

Artificial sediments were formulated for each test (particle size and organic matter content) to maintain consistency and control (Reish and Carr, 1978; Suedel and Roberts, 1994). The materials and procedure were similar to those described by Hill *et al.* (1993). Washed builders' sand and alluvial clay were used in five different proportions for the particle size test. The sand was sieved to exclude particles larger than 2 mm and smaller than 63 μm. The different mixes (by dry weight) were: 100% sand; 90% sand 10% clay, 80% sand 20% clay, 50% sand, 50% clay, 100% clay. Batches were mixed together with an industrial food mixer after drying for 3 days. Each replicate dish received 15 g of the sediment and 100 ml or 80 ml seawater according to the dish type. A disc of polythene was placed over the top of the sediment whilst adding water to minimize disturbance. All bowls were then conditioned for one week to allow a bacterial flora to develop, before introducing the test organisms. The pH of each replicate was recorded at the beginning and end of the week. No large differences in pH were found; the lowest mean reading of pH 7.48 (n=7) was found in the 100% clay group.

For the organic matter tests, an inorganic mix of 85% sand, 15% clay formed the basis of the sediments. Organic matter in the form of sieved moss peat was used in the following dry weight proportions: 0% (control), 5%, 10%, 15% and 20%. These percentages were used to bracket the range of organic contents that are found in environmental samples (Bearman, 1983). Materials were mixed dry at first, but subsequently required moistening to prevent the peat from separating when water was added to the replicates. The pH in each replicate was then measured. The lowest pH levels were found in the 20% organic matter replicates where the average value was 3.58 (n=14). Such levels would be detrimental to the organisms. At the 5% peat content the average pH was 5.6 (n=14). In the 0% replicates an average of pH 7.49 (n=14) was found. Culture water was recorded at pH 8.06 (n=3). In all but the 0% replicates the pH had to be adjusted to avoid damaging the organisms. This was achieved by gradual additions of calcium carbonate ($CaCO_3$). Although Suedel and Roberts (1994) used $CaCO_3$, its use was regretted in this study. Not only did it take time for the neutralization of the acids to occur, it also took up to 2 g of $CaCO_3$ to achieve the desired pH level ($\geq$ pH 7.5). The amount of $CaCO_3$ required in each replicate did not appear to be proportional to the amount of peat present. $CaCO_3$ is also insoluble and in such quantities it changed the characteristics of the sediment. A thin film

Another bioassay was performed with sediments provided by WRc. These sediments originated from a contaminated Dutch harbour and are described in the main WRc report. *O. hartmanni* was the test species used in this assay.

2.7 Extracting *Ophryotrocha* from sediments

All three species of *Ophryotrocha* used in this study produce mucus whilst living in sediment. This means that the egg masses, larvae and adults are often fixed to the sediment particles or the replicate bowl. Sampling was performed by staining with Rose-Bengal, fixing with 4% formaldehyde and then sieving gently. Two sieves were used: a 250 μm mesh to collect the larger particles, egg capsules and adults, and a 63 μm mesh to collect all eggs and larvae. To prevent a variable number of coelomic eggs from spilling into the sample should the adults burst, adults (and egg capsules of *O. diadema*) were separated from the eggs and larvae. Numbers of larvae and eggs exceeded 1000 in most replicates. To limit the time allocated to counting, sub-sampling was conducted. The contents of the sieves were made up to 140 ml and three sub-samples were extracted from each replicate by agitation of the organisms in a suitable beaker. In the case of the whole sediments, the volume was increased to 280 ml to dilute the greater quantity of fine sediment particles. Using a pipette, 10 ml of the mixture was withdrawn for each sub-sample. Larvae and eggs were then counted using a counting crucible under a 10-30X binocular microscope.

2.8 Water-only acute toxicity tests with *O. hartmanni*

Because of a lack of data on the relative sensitivity to toxicants of the native species of *Ophryotrocha*, we decided to run some acute lethal toxicity tests with common pollutants. Two heavy metals, cadmium and zinc, were selected. Adult *O. hartmanni* were exposed to 0, 0.05, 0.2, 1.0, 5.0 and 10.0 mg Cu l^{-1} as Analar $CuCl_2$, and the same concentrations of zinc as Analar $ZnCl_2$. Five adults were exposed in each of two replicates at each concentration, and survival observed at 1.5, 3, 6, 24, 48, 72, 96, 120 and 144-h. The 96-h LC50 was estimated by the binomial method (Newman, 1996). A No Effect Concentration was also estimated using the proportional hazards model implemented within the Debtox statistical package (Kooijman and Bedaux, 1996).

tests (ANOVA) were used to assess differences in reproductive output between the treatments. This test can be used even when the distribution of data is slightly non-normal, and when variances are heterogeneous (Zar, 1984; Day and Quinn, 1989). When the Shapiro-Wilk test showed a highly significant departure from normality ($p<0.01$), the Kruskal-Wallis non-parametric ANOVA was used. For similar reasons the Bartlett-Box test was not used if p was less than 0.05. The Tukey multiple range test was used after either type of ANOVA to determine which pairs of treatments differed significantly.

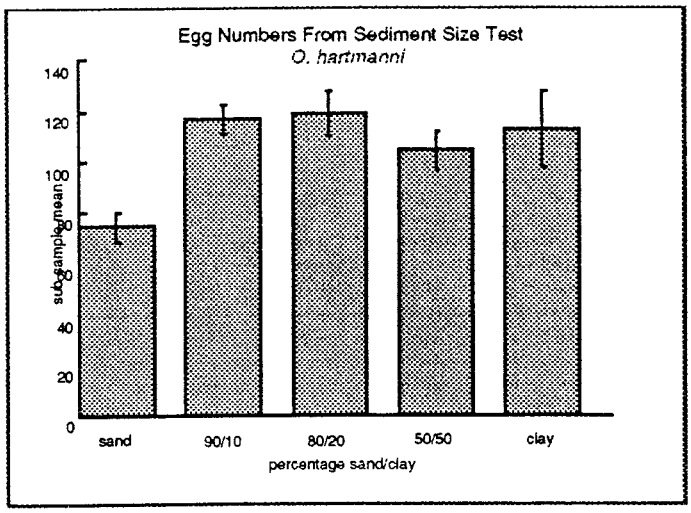
Two measures of reproductive output were measured: number of larvae and number of eggs. The precision and power of the test (Giesy and Hoke, 1989; Reynoldson *et al.*, 1991) were calculated by Pearson and Hartley's method (1951, cited in Zar, 1984). The minimum replicate number required in a test was calculated with an alpha error of 0.05 and beta error of 0.20, based on a minimum specified difference of 30% of the mean. This type of power analysis is important when any statistical analysis is performed (Giesy and Hoke, 1989; Reynoldson *et al.*, 1991), and should be essential when developing new methods.

3.2.1 Sediment particle size test for O. hartmanni

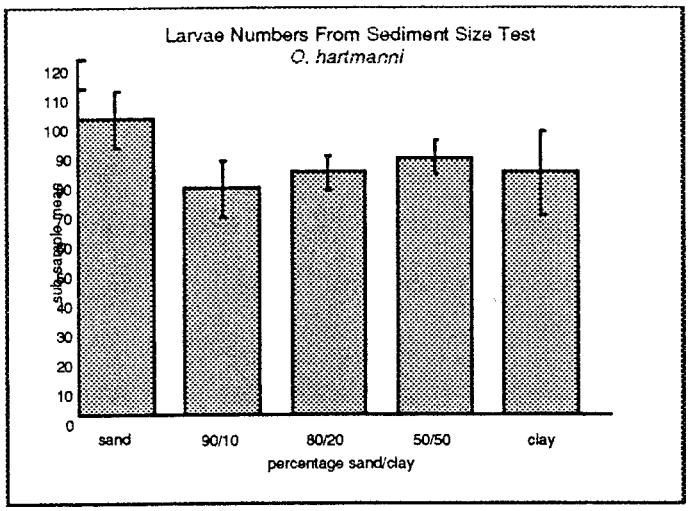
The results from this experiment are shown in Figure 1. The distribution of data was found to depart from normality for only the egg data in one treatment. The p-value was 0.019, and therefore deemed not to be too great a departure from normality to warrant use of the Kruskal-Wallis test. The Bartlett-Box test indicated homogeneity of variances (Table 3).

Variable	Shapiro-Wilk P-value	Bartlett-Box P-value	ANOVA p-value	Rep. No at 30% mean	Differences between groups Tukey HSD
Eggs	0.190-0.596	0.193	0.0082	14	100%S differs to 90%S, 80%S and 100%C
Larvae	0.312-0.680	0.222	0.4304	24	no differences
Eggs + Larvae	0.561-0.866	0.442	0.5462	7	no differences

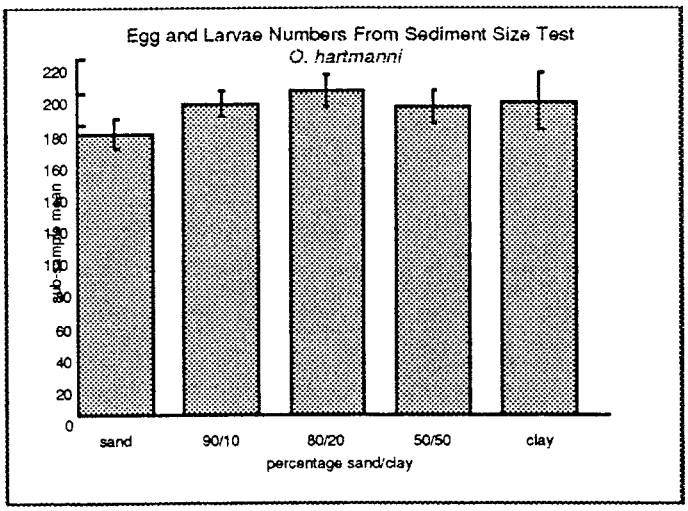
Table 3 Summary statistics of particle size test with *O. hartmanni*.



Graph 1a

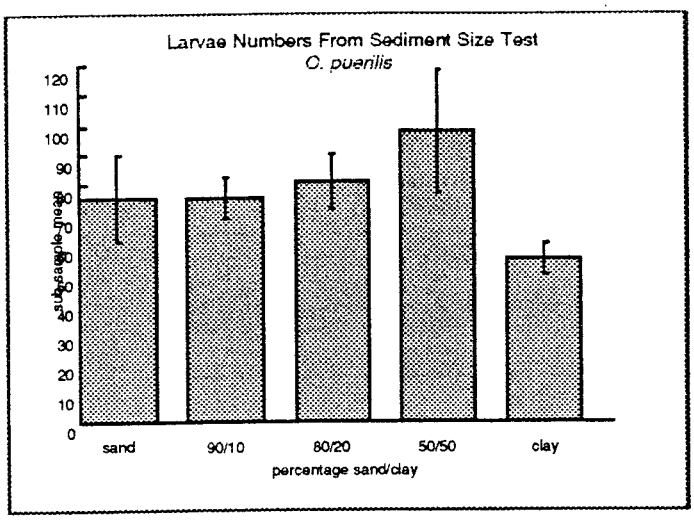


Graph 1b

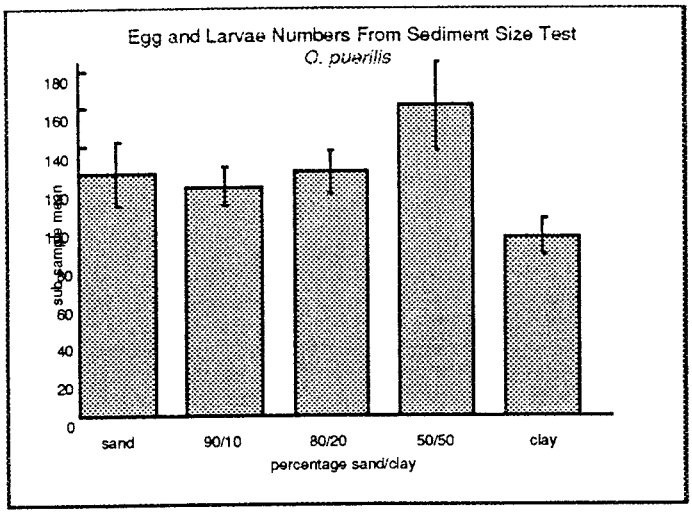


Graph 1c

Figure 1 The effect of particle size on reproductive output in *O. hartmanni*. Error bars denote standard error. Graph a represents the number of eggs, Graph b the mean numbers of larvae in the sub-samples and Graph c the mean number of eggs plus larvae in the sub-samples.



Graph 2b



Graph 2c

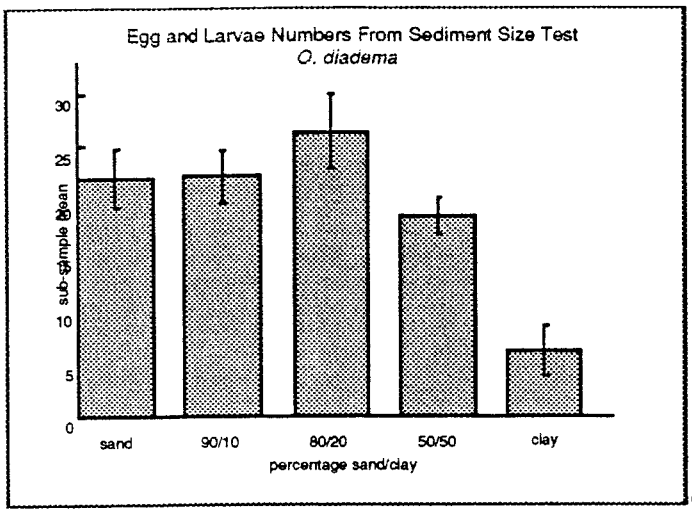
Figure 2 The effect of particle size on reproduction in *O. p. siberti*. Error bars indicate standard error. Reduced output is apparent in the 100% clay treatment, but it is only significant when compared to the 50% sand 50% clay.

3.2.3 Sediment particle size test for O. diadema

Normality in egg frequency distribution was deviated from in three of the five treatments. The p-values were just outside the specified significance levels, so ANOVA was used. The Bartlett-Box test was not used because of the extreme loss of reliability when the data are non-normal (Zar, 1984).

Variable	Shapiro-Wilk P-value	Bartlett-Box P-value	ANOVA P-value	Rep. No at 30% mean	Differences between groups Tukey HSD
Eggs	0.007-0.396	not used	0.195	695	no difference
Larvae	0.011-0.916	0.7	0.0006	54	100%C differs to all others
Eggs + Larvae	0.017-0.729	0.091	0.001	52	100%C differs to all others

Table 5 Summary statistics of sediment size test with *O. diadema*.



Graph 3c

Figure 3 The effect of particle size on reproductive output in *O. diadema*. Error bars are represented by standard error. Note the large error bars in Graph a.

3.2.4 Organic matter test for O. p. siberti

All of the data sets deviated substantially from normal (Table 6). Logging and square-rooting the data did not improve this. Therefore, the Bartlett-Box or power test could not be used (Zar, 1984). In this instance, all three data sets were analyzed using the Kruskal-Wallis test. The coefficients of variation were lower in the sediment size test. Although this is not as useful as the power test, it does provide a simple comparison of variabilities. Control replicates contained an average of 9.4 adult worms, and larvae and eggs were found in each.

Variable	Shapiro-Wilk P-value	Kruskal-Wallis P-value	Differences between groups Tukey HSD
Eggs	0.001-0.469	0.0150	0% is different to 15%
Larvae	0.007-0.402	0.0001	0% is different to all others
Eggs + Larvae	0.009-0.273	0.0022	0% is different to all others

Table 6 Summary analysis for organic matter content test with *O. p. siberti*. Control has greater output than other categories.

All three data sets indicate that *O. p. siberti* suffers a highly significant decrease in reproductive output when exposed to the treatments containing organic matter (Figure 4).

Because the proportion of eggs to larvae is not consistent between the control and the other treatments, it is likely that larval mortality, as well as a reduction in egg production has occurred. Whether this is a reaction to the peat or the $CaCO_3$ is not known. The quantities of $CaCO_3$ added were very similar in all treatments and not proportional. However, acidity was proportional and the bi-product of the reaction (CO_2) may well be a factor in the reduced reproduction. There did not appear to be a proportional increase in the amount of bacteria in treatments, although this cannot be ruled out.

Unfortunately, the culture of *O. hartmanni* did not contain sufficient organisms to include this species in the organic matter tests. This omission should be rectified in any future work.

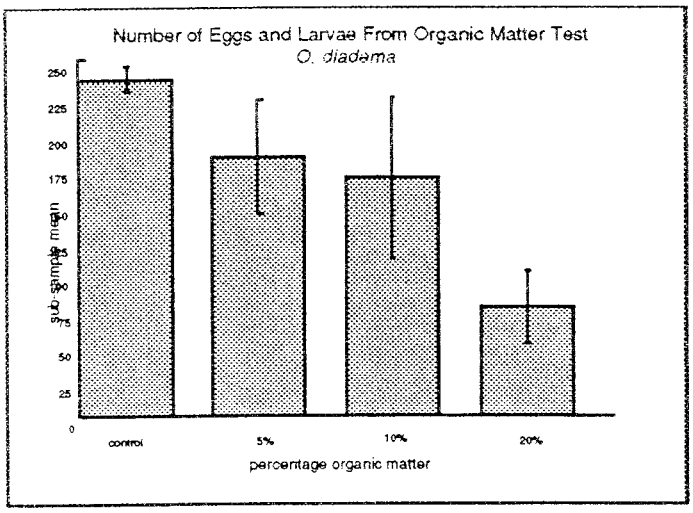
3.2.5 Organic matter test for O. diadema

Table 7 provides summary statistics for the organic matter test with *O. diadema,* and the corresponding graphs are shown in Figure 5. Many of the test replicates failed in this test. Only one replicate of the 15% organic matter content had adults and eggs when sampled, hence this treatment has been excluded from the analysis. The Shapiro-Wilk test revealed large departures from normality in all groups. When square-rooted however, the data fell within our requirements. The unbalanced replicate numbers prevented the use of the Bartlett-Box test. Adult survival in the control was 84%, and each replicate contained young. The test can therefore be accepted as valid.

Variable	Shapiro-Wilk (on groups)	ANOVA P-value	Rep. No at 30% mean	Differences between groups Tukey HSD
Eggs	0.064	0.0082	17	20% differs from 0%
Larvae	0.0097	0.0001	62	0% differs from 5%, 10%, 20%
Eggs + Larvae	0.046	0.0063	17	20% differs from 0%

Table 7 Summary analysis of organic matter content test with *O. diadema.*

Replicate number from power analysis was found to be much lower than the previous test with *O. diadema.* and there was a response pattern similar to the *O. p. siberti* organic matter test. There was a reduction in reproductive output from the 0% organic matter and the other treatments. Significant differences existed between 0% and 20% for egg number,



Graph 5c

Figure 5 Reproductive output of *O. diadema* in the organic matter content test. Error bars denote standard error. Graph a shows results using egg count, Graph b shows the larvae number and Graph c combine egg and larvae numbers.

3.3 Bioassays with River Tyne sediments and *O. p. siberti*

Only four replicates were used with this experiment instead of the seven used in the previous experiments. The three data sets (eggs, larvae, and eggs and larvae) were not found to depart considerably from normal. However, a significant difference was found for two sub-samples and so a Bartlett-Box F-test was not used due to the low reliability when departures from normality are encountered (Zar, 1984).

Variable	Shapiro-Wilk range of P-values.	ANOVA P-value
Eggs	0.0829-0.9801	0.3566
Larvae	0.0092-0.957	0.0892
Eggs + Larvae	0.0683-0.7015	0.2780

Table 8 Summary statistics from a field sediment bioassay with sediments collected from the River Tyne using *O. p. siberti*.

Although no significant differences were found between the stations from which sediment was collected, there was a reduction in output of larvae between St Anthony's and Rohn at the confidence level of 91% (see Figure 6b). The standard error is high for all three endpoints and so this test may not be very reliable. Also, the sediment that was used as the control (River Avon) produced one of the lowest outputs from all three endpoints. We are

3.4 Bioassays with Dutch sediments and *O. hartmanni*

The results from the bioassay of Dutch sediments with *O. hartmanni* showed that these sediments were not lethally toxic to adults. Survival was ≥ 80% in all but two replicates (it was 60% in one replicate from station 2 and 70% in one replicate from station 3).

There were no significant differences between the sediments in either egg number (ANOVA, $F_{5,24} = 0.916$, $p = 0.49$) or egg + larvae number ($F_{5,24} = 1.026$, $p = 0.42$). However, there was a significant difference in the number of larvae produced ($F_{5,24} = 16.46$, $p < 0.001$), because larval production in all of the Dutch samples was lower than in the control (Tukey test, $p < 0.05$).

3.4 Water-only tests with *O.hartmanni*

The binomial 96-h LC50 values for copper and zinc were 0.4 mg Cu l^{-1} (95% C.I. 0.2-1.0 mg l^{-1}) and 2.5 mg Zn l^{-1} (95% C.I. 1.0-5.0 mg l^{-1}). The No Effect Concentrations estimated by the proportional hazards model were 0.48 mg Cu l^{-1} (95% C.I. 0.1-0.7 mg l^{-1}) and 1.98 mg Zn l^{-1} (95% C.I. 0.7-3.9 mg l^{-1}).

1993). If the autecology of an organism is highly specialized for existence in a particular sediment type, it will be of use only when sediments of that character needed assessing.

Results from the particle size tests show that all three species are equally suitable for bioassays with sediments of different particle sizes. *O. hartmanni* has a potential preference for whole sand, whilst *O. p. siberti* and *O. diadema* prefer sediments that contain some sand. The organic matter tests for *O. p. siberti* and *O. diadema* show a negative relationship for organic matter content (peat). Since most sediments within coastal waters seldom rise above 10% in organic matter content (Bearman, 1989) this may not be an important factor. The method used to test organic matter tolerance may be called into question here. Walsh *et al.* (1992) note that caution must be exercised when formulating sediments, as they can be much too simplistic when compared to field sediments. With the benefit of hindsight, the artificial sediment used in this study was probably too simple. Methods used by Walsh *et al.* (1992) and Suedel and Roberts (1994) are more complex than those suggested by Hill *et al.* (1993), and therefore may be more realistic. Alternatively Chandler (1986) provides a method for cleaning field sediments that also reduce the chances of bacterial blooms. Chandler's method, chemical analysis, and serial dilutions of the clean sediments may be a better way to assess an organism's response to organic matter content.

4.3 Precision of results

Variability in the data obtained from a bioassay must be low enough to allow statistical tests to detect differences. The power calculations show that as many as 695 replicates are needed to achieve a power of 80% in one of the tests reported in this study. Conversely, using 7 replicates in the same tests, the null hypothesis would be accepted when it should be rejected more than 65% of the time. Tests with *O. hartmanni* had the greatest power, which is indicated by the low number of replicates required to achieve 80% power. Variation in the *O. hartmanni* endpoints was probably lower because these animals are sequential hermaphrodites. Once they are female they do not have the ability to produce sperm. The other two species have the option to partition resources between sperm and eggs once they have achieved female size (Åkesson, 1973; and Tennant, 1984). This can result in different numbers of egg-producing females in each replicate. The results for *O.*

This total is the minimum that must be allotted to each test. Cleaning and preparation of equipment is excluded. Bench space will also be required for the duration of the test. This will be minimal for the preparation and culture (approx. $0.5m^2$), but will increase to $1m^2$ for the experiment, which lasts for 35 days.

Hooftman and Vink (1989) estimated that a single water-only bioassay using *O. diadema* should take up to 200 hours. The reason for the discrepancy between their estimate and the one found in the present study is that Hooftman and Vink constantly monitored reproductive output throughout their tests.

4.5 Choice of Organisms for Further Study

Monk (1983) suggests that no single species can be universally applicable in ecotoxicological tests. Indeed, many organisms must be made available to ensure that all aspects of ecotoxicology can be addressed. However, to have many organisms of the same genus would result in many other taxa being under-represented. For this reason a choice must be made between the three species studied here.

The use of *O. diadema* are more time consuming and they are sampled less precisely using the method followed in this study. For these reasons *O. diadema* should be rejected in favour of one of the other two species. Also, *O. diadema* are found along the Pacific coast of the United States and are not native to the UK.

Slightly more information has been obtained on *O. p. siberti* than on *O. hartmanni* in this study. However, *O. hartmanni* endpoints have proved to be less variable and may be more suitable for tests using this method.

5. CONCLUSIONS AND RECOMMENDATIONS

1. All three species of *Ophryotrocha* have short life-cycles compared to many organisms currently in use and are available in laboratory culture for toxicity tests and bioassays throughout the year.

6. ACKNOWLEDGEMENTS

We should like to thank the following people: Professor Bertil Åkesson for supplying the first batch of worms, and for his support and provision of information during the study. Similarly, Dr. John Knowles of MAFF for information and for worms. Dr Yvonne Allen of MAFF for the supply of field sediment from the Tyne, and Rachel Fleming of WRc for providing the Dutch sediments. Neil Somerville for his help in counting animals and Ray Jalland for help with setting-up experiments and checking water quality. Tim Allen for running the water-only toxicity tests. Professor Geoff Moore and Belinda Luke for their advice on sub-sampling techniques. Finally, Professor John Dodge for the identification of the algae found in the cultures.

Burgess, R.M., Schweitzer, K.A, McKinney, R.A. and Phelps, D.K. (1993) Contaminated marine sediments: water column and interstitial toxic effects. *Environmental Toxicology and Chemistry*, **12**, 127-138.

Cairns, J.C. Jnr, (1986) What is meant by validation of prediction based on laboratory toxicity tests? *Hydrobiologia*, **137**, 271-278.

Calow, P (1989) The choice and implementation of environmental bioassays. *Hydrobiologia*, **188/189**, 61-64.

Chandler, G.T. (1986) High-density culture of meiobenthic harpacticoid copepods within a muddy sediment substrate. *Canadian Journal of Fisheries and Aquatic Science*, **43**, 53-59.

Chandler, G.T. (1990) Effects of sediment-bound residues of the pyrethroid insecticide fenvalerate on survival and reproduction of meiobenthic copepods. *Marine Environmental Research*, **29**, 65-76.

Chandler, G.T., Coull, B.C. and Davis, J.C. (1994) Sediment and aqueous-phase fenvalerate effects on meiobenthos: implications for sediment quality criteria development. *Marine Environmental Research*, **37**, 313-327.

Chandler, G.T. and Scott, G.I. (1991) Effects of sediment-bound endosulphan on survival, reproduction and larval settlement of meiobenthic polychaetes and copepods. *Environmental Toxicology and Chemistry*, **10**, 375-382.

Crane, M., Everts, J., van de Guchte, K., Heimbach, F., Hill, I.R., Matthiessen, P. and Stronkhorst, J. Research needs in sediment bioassay and toxicity testing. *Ecovision World Monograph Series (in press)*.

Day, K.E., Dutka, B.J., Kwan, K.K., Batista, N., Reynoldson, T.B. and Metcalfe-Smith, J.L. (1995) Correlations between solid-phase microbial screening assays, whole-sediment

Gray, J.S. (1992) Biological and ecological effects of marine pollutants and their detection. *Marine Pollution Bulletin*, **25**, 48-50.

Hardy, B.L.S. (1978) A method for rearing sand-dwelling harpacticoid copepods in experimental conditions. *Journal of Experimental Marine Biology and Ecology*, **34**, 143-149.

Hermsen, W., Sims, I. and Crane, M. (1994) The bioavailability and toxicity to *Mytilus edulis* L. of two organochlorine pesticides adsorbed to suspended solids. *Marine Environmental Research* **38,** 61-69.

Hill, I.R., Matthiessen, P. and Heimbach, F. (1993). *Guidance Document on Sediment Toxicity Tests and Bioassays for Freshwater and Marine Environments.* From the "Workshop On Sediment Toxicity Assessment". SETAC Europe, Brussels.

Hooftman, R.N. and Vink, G.J. (1980) The determination of toxic effects of pollutants with the marine polychaete worm *Ophryotrocha diadema. Ecotoxicology and Environmental Safety*, **4**, 252-262.

Klockner, K. (1979) Uptake and accumulation of cadmium by *Ophryotrocha diadema* (Polychaeta). *Marine Ecology-Progress Series*, **1**, 71-76.

Knezovich, J.P., Harrison, F.L. and Wilhelm, R.G. (1987) The bioavailability of sediment-sorbed organic chemicals: a review. *Water, Air and Soil Pollution*, **32**, 233-245.

Knowles, J.F. and Greenwood, L.N. (1994) The effects of chronic radiation on the reproductive performance of *Ophryotrocha diadema* (Polychaeta, Dorvilleidae). *Marine Environmental Research*, **38**, 207-224.

Knox, G.A. (1977) The role of polychaetes in benthic soft-bottom communities. In, Reish D.J. and Fauchald K. (eds). *Essays on Polychaetous Annelids in Memory of Dr. Olga Hartman*, pp. 547-604. Allan Hancock Foundation, University of Southern California, Los Angeles.

Reish, D.J. and Carr, R.S. (1978) The effects of heavy metals on the survival, reproduction, development and life cycles of two species of polychaetous annelid. *Marine Pollution Bulletin*, **9**, 24-27.

Reynoldson, T.B. (1994) A field test of a sediment bioassay with the oligochaete worm *Tubifex tubifex* (Müller, 1774). *Hydrobiologia*, **278**, 223-230.

Reynoldson, T.B., Thompson, S.P. and Bamsay, J.L. (1991) A sediment bioassay with the tubificid, oligochaete worm *Tubifex tubifex*. *Environmental Toxicology and Chemistry*, **10**, 1061-1072.

Roddie, B., Kedwards, T., Ashby-Crane, R. and Crane, M. (1994) The toxicity to *Corophium volutator* (Pallas) of beach sand contaminated by a spillage of crude oil. *Chemosphere* **29**, 719-727.

Selley, R.C. (1982) *An Introduction to Sedimentology*, 2nd Edition. Academic Press, London.

Shapiro, S.S., Wilk, M.B. and Chen, H.J. (1968) A comparative study of various tests for normality. *American Statistical Association Journal*, **63**, 1343-1372.

Skelton, P. (1993) *Evolution, A Biological and Palaeontological Approach.* Addison-Wesley, Wokingham, UK.

Smith, D.P., Kennedy, J.H. and Dickson, K.L. (1991) An evaluation of a naidid oligochaete as a toxicity test organism. *Environmental Toxicology and Chemistry*, 10, 1459-1465.

Suedel, B.C. and Roberts, J.H. (1994) Development of formulated reference sediments for freshwater and estuarine sediment testing. *Environmental Toxicology and Chemistry*, **13**, 1163-1175.

APPENDIX B CHRONIC PROTOCOLS

Institute for Inland Water Management and Waste Water Treatment (RIZA)

PROTOCOL FOR TESTING OF FIELD SEDIMENT IN CHRONIC BIOASSAYS WITH THE FRESHWATER DIPTERAN *CHIRONOMUS RIPARIUS*

second version: JANUARI 1996

Working document 95.192X

Authors E.M.M. Grootelaar, J.L. Maas, L.C.M. Kerkum-Mank and C. van de Guchte
RIZA, Postbox 17, 8200 AA Lelystad, The Netherlands
Division for Ecotoxicology report nr. WSCE 96-01

1. Introduction

This protocol describes a method for measuring the chronic effects of chironomids in field freshwater sediments.

This protocol is applicable to most field sediments with the exception of for example sandy sediments, which are unsuitable as habitat for the chironomid *Chironomus riparius*. A reference sediment is used to check the quality of the testorganisms.

In this Toxicity test larval instars of chironomid larvae are exposed, starting with egg packets. The first larval instars are the most sensitive and dominant part of the life cycle of chironomids. Chironomids are present in the sediment during their larval and the pupal stages. This protocol is intended to define lethal and sublethal effects of contaminants on the specific stages of the species tested.

Prior to the publication of this protocol, experience with this larvae test has mostly been with the freshwater chironomid *Chironomus riparius* (Diptera: Chironomidae). More detailed information on this species is therefore given in Appendix 1.

2. Principle of the test

The egg packets of chironomids are exposed to the overlying water of the sediment tested. The test is a semi-static procedure. The test is started by placing egg packets in test vessels with overlying water only. After one week hatched larvae are replaced in a testsystem containing whole sediment and overlying water. The test ends just before the pupa stage. The endpoints are: survival, development (different instars), growth (dry weight of L4 larvae) and abnormal behaviour (such as avoidance of the sediment).
Lethal and sublethal effects are assessed and compared with the control values of a reference sediment.

3. Test animals and equipment

Species : *Chironomus riparius*
Source : laboratory culture, maintained as described by Grootelaar et al. (1993).

The four different larval instars are separated on the basis of headcapsule width (Appendix 1) using a lightmicroscope. Egg packets are used as test organisms.
The sensitivity of the cultured midge larvae (second instar) should be determined by an acute test in water with potassium dichromate ($K_2Cr_2O_7$) or 3,4-dichloroaniline as reference compound (Appendix 3).

5. Sediment and water

Any reconstituted water in which the test species meet the validity criteria for survival and development is suitable as dilution water.
Any relatively clean sediment in which the test species shows control survival and development is suitable as a reference sediment. A characterization of the sediment is carried out including the following parameters: the level of dry matter, particle size distribution, (an)organic carbon, N(Kj-N), P(P_2O_5) and oil (Appendix 4).

6. Pre-treatment of sediment

Sample ca. 30 litre of sediment (upper layer of 10-15 cm) by an Ekman-Birge sampler or box-corer at the field locations and transfer the sediment in buckets of 10 litre. Store the buckets with sediment at 5ºC until use (max. 3 months after sampling). Small pieces of stone, valves and wood in the sediment should be removed. Any microfauna present should be removed from the sediment by sieving through a 500 µm sieve. Homogenise the sediment with the overlying water and store it at 5ºC during one day to settle the sediment. After one day, decant the overlying water until the sediment is just wet. Determine the level of dry matter of the sediment and the other sediment parameters. Determine the organic and anorganic micro-contaminants as well. Store the sediment at 5ºC until use.

7. Method

The test starts by placing egg packets into **two** test vessels per sediment with overlying water only. After one week the hatched larvae are transferred in **four** test vessels per sediment containing whole sediment and overlying water.
The test should be carried out at a temperature of 20 $\pm$ 1ºC with a photoperiod of 16 hours light and 8 hours dark and a light intensity of 2000-2500 lux. During the test period the overlying water is to be aerated when necessary and the animals are fed.

For each sediment a separate overlying water (t=0-7d) and a sediment-water-system (t=7-28d) should be prepared.

1. Add 100 ml homogenized pretreated field sediment to 400 ml dilution water (volume ratio of 1:4) in a 800 ml bottle. Shake or roll the sediment-mixtures for 20 hours, to reach an equilibrium between the sediment- and the waterphase.

2. To obtain the overlying water, decant 200 ml of the sediment-suspension into another vessel. Leave these vessels for three days at 20ºC in order to allow the sediment to settle. Store the bottles with the remaining 300 ml suspension at 5ºC until use (7.7).

3. After the sediment has settled, decant the overlying water and place 50 ml overlying water in each of the duplicate test vessels.
If the overlying water is turbid, centrifuge it during 30 min. at least 2500 x g.

4. Measure dissolved oxygen, pH, temperature and NO_3/NO_2 –levels (testkit) in each test vessels. If necessary aerate moderately for ca. 10 minutes. Measure salinity at the beginning of the test. Measure temperature in the test media at least three times a week. Measurements

A schematic description of the preparation of the difference test-systems **per field sediment**:

t=0 day:
2 testvessels with overlying water only → 3 half egg packets per test vessel

t=7 days:
4 testvessels with whole sediment and overlying water → 25 larvae per test vessel

8. Statistics

Lethal and sublethal effects (larval development and dry weight) are assessed and compared with reference values.

Survival:
Count the number of living larvae, pupea and midges and calculate the mean survival of the four test vessels per sediment.

Larval development:
Count the number of fourth instars (alive) together with the pupae and midges (dead or alive) and calculate the mean larval development. Use a binocular to determine the larval instars.

Dry weight:
Calculate the dry weight of the living fourth instar per testvessel (mg per L4 larva).

Variations within each set of replicates are analysed using analysis of variance or contingency table procedures. In order to make a multiple comparison between the results at the individual sediments and those for the reference the NOVA-Dunnetts-test or bonferroni (Sokal and Rohlf, 1981: Dunnetts, 1955 and 1964) is useful. You can find this test in the program Toxstat (version 3.3; Gulley et al., 1988).

Appendix 1

Information on the freshwater chironomid *Chironomus riparius*

Ecologically, chironomids form an important group of organisms. A large part of their life cycle, the larval stage, is spent in sediment; during the egg stage and the adult midge stage they live in the water and in the air respectively. They are a part of different food chains : egg, larvae and pupae comprize an important source of food for aquatic organisms (e.g. fish) and midges act as food for terrestrial organisms (e.g. birds and bats).

Various chironomids can be found in most sediments except sand. *Chrionomus plumosus* is highly abundant in large rivers and lakes. *Chironomus riparius* is found in shallow pools and lakes. *Chironomus riparius* is a deposit feeder, the larvae feed on small silt particles with a diameter of 5 – 35 µm (silt), or on microdetritus (Rasmussen, 1984).

The appearance of different larval stages is dependent upon the season. Only fourth stage larvae survive during winter in the sediment. The first midges emerge in spring and many young larvae (second stage) are subsequently found. In summer, all larval stages can be found. In general there are two periods of swarming, in late spring and at the beginning of autumn.

The entire life cycle lasts ca. 4 weeks at a temperature of 20ºC (table 1). The fertilized egg packets hatch 2 to 3 days after having been laid in water. The larvae undergo four stages, distinguishable by differences in head capsule width (Table 1). Second, third and fourth stage larvae nest in selfmade tubes in the sediment. The larval stages last three weeks. After the fourth stage, the larvae change into pupae. Immediately after the pupa moves to the surface of the water the midge emerges from the pupal skin. The emergence of male midges starts earlier (2 days) than that of the females. Females are a little large and darker than males. Males have two hairy antennae on their heads.
Mating occurs during swarming. After 1 – 2 days the female midge lays an egg packet (ca. 600 eggs); 2 to 3 days later a second, but smaller, egg packet (ca. 200 eggs) can be produced. Midges have a live time from four to eight days. *Chironomus riparius* is easily reared in the laboratory. *Chironomus riparius* only needs a small flying height in order to mate (ca. 5 cm) in contrast to other chironomids like *Chironomus plumosus*.

Appendix 3

Acute toxicity test to control the sensitivity of the cultured midge *Chironomus riparius*

Method

The test should be carried out at a temperature of 20 ± 1ºC with a photoperiod of 16 hours light and 8 hours dark and a light intensity of 2000-2500 lux. The survival in the control must be at least 90% after 96 hours.

1. Prepare a concentration range of potassiumdichromate ($K_2Cr_2O_7$) or 3,4 – dichloraniline in 50 ml DSW in 100 ml glass test vessels. The concentration range of $K_2Cr_2O_7$ should be, control, 10, 18, 32, 56 and 100 mg/l. The concentration range of 3,4 – DCA should be, control, control and Tertiair Butyl Alcohol (TBA), 1.8, 3.2, 5.6, 10.0 and 18.0 mg/l.

2. Use duplicate test vessels. Add 25 second instar larvae in each test vessel.

3. Feed the larvae at t=0 and t=48h with 200 µl of a 2% (2g/100 mls) foodsuspension (Trouvit).

4. Observe the behaviour and survival of the midge larvae every day.

5. Record at t=0 and t=96h Dissolved Oxygen, pH and NO_3NO_2 – and NH_4 –levels (testkit) in each test vessel. Measure these conditions also if high mortality is observed in a concentration.

6. Do not aerate during the test.

Statistics

Calculate the LC50 (EC50) and their 95% confidence interval by means of a parametric model developed by Kooijman (1981). This is a maximum likelihood estimation with probits of the number of deaths as function of the logarithms of the corresponding concentrations.

EC50: The concentration at which a 50% effect occurs.
LC50: The concentration at which 50% mortality occurs.

Lethal and sublethal effects are assessed and compared with control values to determine the lowest observed effect concentration (LOEC) and hence the no observed effect concentration (NOEC).

The ranges of the reference compounds potassium dichromate ($K_2Cr_2O_7$) and 3,4 – dichloroaniline for an acceptable sensitivity of *Chironomus riparius* in a laboratory culture are given below.

Appendix 4

Reference sediment and standard water

Characterization of reference sediments used for chironomid tests.

Chemical parameters	sed. 1	sed. 2	sed. 3
particle size (%)			
< 63 (µm)	61.6	25.8	54.9
< 50 (µm)	56.3	22.5	50.6
< 16 (µm)	28.1	8.6	23.1
< 10 (µm)	20.1	5.5	15.7
< 2 (µm)	5.5	1.0	4.2
dry matter (%)	36	17	
org. C.* (%)	3.38	6.84	2.99
anorg. C.* (%)	1.42	4.51	2.44
N (Kj-N)* (g/kg)	3.62	6.28	2.56
P (P_2O_5)* (g/kg)	0.97	1.64	6.70
oil (mg/kg)	78.0	94.2	373.9

* as dry matter content of sediment

INFORMATION ON THE TEST SUBSTANCE

5. Useful information about the test substance includes: structural formula, purity of the test substance, stability and biodegrability in water, light and sediment, water solubility and n-Octanol/water partition coefficient.

DESCRIPTION OF THE METHOD

Collection of control sediment and animals

6. Control sediment and animals should be collected from the same, clean intertidal mudflat, estuary with a large population of the test species.
The sediment is wet-sieved through 500-um mesh sieve in the field to remove small amphipods and other macrobenthos. The amphipods retained on the sieve are poured into a clean polyethylene bucket half-filled with natural seawater and transported to the laboratory as soon as possible.

Characterization and storage

7. Prior to testing, a sample of the control sediment is analysed for organic carbon content and particle size distribution.
Further characterization might include: total metals, total chlorinated organic compounds, polycyclic organic hydrocarbons, etc. (1)

8. The control sediment must be stored in glass or rigid plastic containers for no longer than two weeks at 4 $\pm$ 2ºC until the start of the test. (2)(3).

Holding of amphipods

9. Animals must be acclimatized for at least 96h to the same salinity, temperature, oxygen concentration, water quality and light conditions as to be used during testing. More time for acclimatization might be necessary (e.g. one week) if the conditions of temperature and salinity are very different than the ones used for testing (i.e. difference in temperature and salinity $\geq$ 5ºC and $\geq$ 5 ppt. respectively). In these cases the acclimatization process should be performed step by step.
In the case of *C. volutator*, the following conditions before and during testing must be maintained (in the overlying water):

Salinity: salinity should be similar to the salinity of the water at the site where sediment and animals were collected (a salinity range of 20 to 30 ppt is advised; however this species is able to survive salinities ranging from 2 to 50 ppt).

Temperature: ranged between 15ºC (minimum) and 18 $\pm$ 1ºC (maximum).

Aeration: at least 60 per cent of air saturation (i.e. at least 6 mg/L at 18ºC and at a salinity of 32 ppt) to be realised by air supply with a Pasteur pipette connected with the air pressure system of the laboratory which ensures constant bubbling.

Water quality: natural sandbed filtered seawater is advised, i.e. water containing suspended particles and organisms smaller than 10 um.

Light: illumination must be uniform and continuous throughout the test period.

PROCEDURE

Set up of an experiment and conditions of exposure

13. 1-L borosilicate glass beakers are filled with approx. 200 mL of test sediment and with approx. 700 mL overlying filtered seawater without resuspending the sediment. In each glass-beaker the air supply is provided by a Pasteur pipette. The vessels are placed in a temperature-controlled waterbath (or room temperature) at a constant temperature ranged between min. 15ºC and max. 18 ± 1ºC for 24h in order to enhance the sediment to settle and to stabilize. After 24h the animals are wet-sieved from the sediment and transferred in a vessel with sea water. The animals are then sized as described in par. 12. and healthy and swimming organisms of the same size are then added (n=20) in each of the 1-L glass beaker.

14. Time of exposure:	28 days
Number of test concentrations:	5
Number or replicates for each treatment:	3 to 5
Number or replicates for the negative control (or solvent control):	6 to 10 (twice the number of replicates in the treatments)
Number of animals per replicate:	20
Endpoints:	- growth (by measuring dry weight) e.g. EC_{20}/EC_{50}, NOEC) - 28-day mortality (LC_{20} or LC_{50} where possible)

15. For the dry weight determination, prior to the start of the experiment 3 to 6 representative subsamples (of 20 animals each) of the selected organisms are taken and transferred into polyethylene vials. The empty vials should have been numerated, dried (at the same temperature and for the same period of time as used for the samples) and weighed before use.
The estimation of the initial weight (at t=0) has only the aim to ascertain whether the type of sediment used is suitable for long-term testing. This values will not be taken into account in the treatment of final results.

16. On the 28th day, at the end of the experiment, the test vessels are taken off by wet-sieving the content through a 500-µm sieve. Animals that do not move even after gentle stimulation with pincers are considered dead.
All survivors of each replicate are counted and placed together in vials which are previously numerated, dried and weighed (as described in par. 15.) so that each vial corresponds to a specific replicate. Vials containing amphipods are dried immediately at 70-100ºC for 48 hours. After drying these are weighed on a balance. The difference between the two weights (vials without and with organisms) equals the weight of the organisms.

17. The effects on dry weight will be estimated by calculating the statistical (significant) differences between the average weight of each treatment group and the controls using the statistical procedure. Analysis of Variance (ANOVA) followed by a multiple comparison test (see par. 24).

Stock solution and spiking method

20. Various methods of spiking are available and described in the literature. The method can be arbitrary chosen by the researcher but must be the same as the one used for the acute tests in order to be able to compare the effects (and the concentration range). This standard operating procedure is based on the 'direct (solvent) spike' method (7). A stock solution is prepared by dissolving the selected substance in an organic solvent (e.g. acetone).

Observation

21. The air supply to the test vessels has to be checked daily to maintain the dissolved oxygen concentration or the aeration bubbling constant.

22. Temperature, dissolved oxygen and salinity in the test vessels are to be recorded at 72h intervals and should stay within the limits mentioned above (see par. 9). Increases in the salinity values due to water evaporation must be adjusted by adding deionised water to each test vessel up the original level. Renewal of overlying water is not necessary during 28-d testing period.

CONDITIONS FOR THE VALIDITY OF THE TEST

23. A test is considered to be valid if the following conditions are fulfilled:

a) all animals have been acclimatized at the test temperature and salinity for at least 96h prior to testing;
b) parameters such as salinity, temperature, dissolved oxygen concentration, are maintained as far as possible constant for the whole period and within the limits mentioned above (see par. 9);
c) mortality in the control(s) does not exceed 20 per cent in case of field-collected animals at the end of the test;
d) mortality in the toxicant treatments does not exceed 50 per cent at the end of the test period.

DATA AND REPORTING

Treatment of results

24. Prior to examinate whether there may be significant lethal or/and sub-lethal effects within treatments, the assumptions of normality and homogeneity of variances of the data are to be checked (with for e.g. the Shapiro-Wilks or Liliefors and the Bartlett's test, respectively). Arc sin square root transformation for mortality data and square root or log transformation for the data concerning growth can be applied. Re-testing for normality and homogeneity of variance has to be repeated. In case that parametric assumptions are still not met, non-parametric tests should be used (i.e. Steel's Many-One Rank Test, Kruskal-Wallis test etc.).

25. If parametric assumptions are met, the various treatments used in the tests (different exposure concentrations) can be compared statistically by a (two-tailed) ANOVA followed by a multiple comparison.
If the F-test with a null hypothesis of no difference among treatments is not statistically significant ($p>0.05$), it can be concluded that the effects observed in each treatment are not large enough to be detected as statistically significant from the control ($p>0.05$).
At the contrary, if the F-test is statistically different, all exposure effects are compared with the control by using a multiple comparison test such as the Williams or the Bonferroni test (8)(9).

- anything unusual or problems (technical or/and biological) encountered during tests;
- discussion of the results.